EXPLORING THE EVOLUTIONARY HISTORY OF CULTIVATED RICE: THE ORIGIN AND EVOLUTION OF FRAGRANCE AND THE GENETIC CONTROL OF BLACK HULL COLORATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Michael James Kovach

May 2010

EXPLORING THE EVOLUTIONARY HISTORY OF CULTIVATED RICE: THE ORIGIN AND EVOLUTION OF FRAGRANCE AND THE GENETIC CONTROL OF BLACK HULL COLORATION

Michael James Kovach, Ph.D.

Cornell University 2010

Cultivated Asian rice traveled a long and complex journey from a low-yielding, weedy grass species to the high-yielding staple crop consumed by billions of people today. This journey, driven by human selection, involved a series of genetic changes that transformed the rice plant in many profound ways. Modern genetics and genomics techniques have made it possible to re-trace the history of rice domestication and evolution. This new knowledge not only renders a clearer picture of the evolutionary paths traveled during rice domestication, but it also provides new insights for plant breeders, who are faced with the challenge of feeding an ever-growing human population. This dissertation examines the current body of knowledge pertaining to rice evolution, and in the process, attempts to improve our understanding of the genetic diversity within *Oryza sativa* and its wild progenitor, *Oryza rufipogon*. An in-depth haplotype analysis is presented to reveal the origin and evolution of fragrance in rice, which remains one of the most important grain quality characteristics from the perspective of both the international export industry and indigenous peoples who have treasured this trait for centuries. The evolutionary history of the major allele responsible for fragrance in most modern varieties, *badh2.1*, is investigated, as well as the origins of several novel fragrance-causing alleles of *BADH2* in unique rice germplasm from across Asia. Also, in an effort to further our understanding of the rice domestication syndrome, a genetic mapping study is presented that identifies the

causal mutation responsible for a fundamental change during rice domestication: the loss of black pigmentation from the outer covering (hull) of the rice seed. A striking genetic phenomenon is revealed, in which the black hull trait is found to be controlled by an epistatic relationship between two physically linked genes.

BIOGRAPHICAL SKETCH

Michael was born November 11, 1983 in Baltimore, Maryland to Gerald and Kathleen Kovach. He grew up in Glen Burnie, MD where he attended Point Pleasant Elementary School until the 5th grade, at which point his family moved to Duncansville, PA. There, Michael attended St. Patrick's parochial school and then Bishop Guilfoyle High School. Michael was an active member of the cross country team, track and field team, and the ski club. He was also an avid home gardener, and spent his summers working for the Baronner Vegetable Farm in Hollidaysburg, PA, inspiring him to pursue a career in the plant sciences. Michael graduated as valedictorian of his senior class, and was accepted into the Schreyer Honors College at the Pennsylvania State University. While pursuing a Bachelors of Science in Horticulture at Penn State, Michael served as President of the Horticulture Club, completed a life sciences internship at the University of Missouri, conducted a research project on wild tomatoes, and had the opportunity to study briefly in Peru and New Zealand. Michael graduated in 2005 as the College of Agriculture Sciences Class Marshall and entered the PhD program in Plant Breeding & Genetics at Cornell University. Michael's experiences at Cornell have included attending a workshop at the International Rice Research Institute in the Philippines, studying international agriculture in southern India, presenting his research at a conference in the U.K., and teaching a molecular breeding course at the Biosciences Eastern and Central Africa (BecA) hub in Nairobi, Kenya. Michael is beginning his plant breeding career as a maize Line Development Breeder for Monsanto Co. in Thomasboro, Illinois.

ACKNOWLEDGMENTS

Foremost and above all, I thank God for providing me with life, a loving family, and the vast opportunities and privileges that have made it possible for me to come this far. My awe at the biological complexities of His creation drives my desire to use crop breeding as a means of maximizing life's potential.

My Mom and Dad, who have sacrificed so much to provide me with an excellent education and who have stayed with me through every fault and tribulation, deserve more acknowledgement than could ever be expressed. Without their loving support, I would never have had access to the resources I needed to pursue this degree, nor the self-confidence to always strive to be the best I can be in all that I do.

To Susan, who took a chance on me, and proceeded to form me into the thinker, the researcher, the plant breeder, and the person I am today. Also to Rebecca and Steve, for helping me to navigate the bumpy road of a PhD career.

I am grateful to the members of the McCouch research program for their valuable critical analysis and helpful comments on my work.

I acknowledge financial support from the Plant Genome Program of the National Science Foundation (Award Numbers #0606461 and #0110004), the EU project META-PHOR (FOOD-CT-2006-036220), the College of Agriculture and Life Sciences, and the Department of Plant Breeding and Genetics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Throughout this dissertation, the nomenclature for genes and alleles will generally follow the revised Gene Nomenclature System for Rice put forth by the Committee on Gene Symbolization, Nomenclature and Linkage, Rice Genetics Cooperative (McCouch & CGSNL, 2008).

The two widely recognized varietal groups of *Oryza sativa*, sometimes referred to as subspecies, are *Indica* and *Japonica*. To avoid confusion between the varietal groups and the five subpopulations of *Oryza sativa*, the following notation will be used throughout this dissertation: When referring to varietal groups, capital letters will be used (ie, *Indica* and *Japonica*). When referring to subpopulations, lowercase letters will be used with the exception of *Group V* (ie, *indica*, *aus*, *temperate japonica*, *tropical japonica*, and *Group V*). The *Group V* subpopulation was formerly referred to as the "*aromatic*" subpopulation, but this designation leads to confusion and will therefore be avoided. Instead, the subpopulation formerly called the "*aromatic*" subpopulation will be referred to by its isozyme classification, *Group V* (Glaszmann, 1987).

Chapter 2 was published as a review in Current Opinion in Plant Biology (Kovach et al., 2008; 11:193). Chapter 3 was published as a review in Trends in Genetics (Kovach et al., 2007; 23:578). Chapter 4 was published as an original research article in Proceedings of the National Academy of Sciences of the USA (Kovach et al., 2009; 106:14444). An expanded study that includes Chapter 5 was published as an original research article in Molecular Breeding (Asante, Kovach et al., 2010; DOI 10.1007/s11032-009-9382-8).

CHAPTER 1:

INTRODUCTION

The Domestication Syndrome in Rice

The transformation of Asian wild rice (*Oryza rufipogon*), a species with high biomass and low grain production, into its cultivated form (*Oryza sativa*), involved a series of profound changes in the rice plant. These changes altered the physiology, architecture, and adaptation of rice, propelling this relatively unmanageable plant to one of the pinnacles of human staple crops. Ultimately, these changes in the rice plant were brought about by ancient humans, who first scavenged the grain for sustenance, and later began to nurture and cultivate the plant. Through centuries of cultivation, humans imposed their desires upon the rice plant to make it more amenable for production and consumption. The selective forces inflicted by humans led to the suite of altered traits we now refer to as the "domestication syndrome", which distinguish *O. sativa* from its *O. rufipogon* ancestor.

O. sativa has become an important source of food and livelihood for nearly half of the human population (Khush, 1997), and has therefore been the focus of intense breeding efforts. Advances in the fields of genetics and genomics have made it possible to probe the history of rice domestication, uncovering the specific genetic mechanisms upon which human selection acted. Many of the genes responsible for the "domestication syndrome" in rice have now been cloned and characterized, allowing us to piece together the complex path traveled by rice as it evolved into its modern form (McCouch et al., 2010).

Some of the traits altered during domestication included dramatic morphological changes that would have made it easier for humans to cultivate and collect the rice

grain. These include a more upright plant habit (Jin et al., 2008; Tan et al., 2008) that lifted the prostrate rice plant off the ground, and altered branching habits (Komatsu et al., 2003; Li et al., 2003a; Li et al., 2003b; Yu et al., 2008) that reduced competition between plants and allowed closer planting. The cultivated species shifted from a perennial habit to an annual habit partially due to the loss of rhizomes (Hu et al., 2003), and this would have encouraged the plant to produce more grain as a means of survival. One of the most essential features of a seed-bearing plant in the wild is ensuring a means of seed dispersal (Gepts, 2004). Upon maturity, wild rice seeds promptly fall from the panicle to propagate the next generation, a feature that would have likely forced humans to harvest the grain prior to full maturity, lest they lose the grain on the ground (Fuller et al., 2007). The loss of seed shattering, meaning the grain remains attached to the panicle at maturity, would have greatly improved the efficiency of harvest for humans (Konishi et al., 2006; Li et al., 2006). Also, while the wild rice plant possesses a hair-like extension of the seed covering (awn) to allow wind and animal dispersal, most *O. sativa* varieties now lack this feature, and QTLs for this trait have been identified (Xiong et al., 1999; Cai et al., 2002; Thomson et al., 2003).

In addition to these conspicuous morphological changes to the rice plant during domestication, other less visible features were altered. In the wild, regular fluctuations in rainfall and temperature and annual variability in climate have encouraged a feature known as seed dormancy. This allows the staggered germination of seeds when conditions are favorable, improving the chances of successful propagation. A loss of this feature would have allowed for synchronous germination, making the rice plant more amenable to agricultural practices (Gu et al., 2004; Gu et al., 2006; Gu et al., 2008). Domestication also brought about a slow shift from the outcrossing habit of *O.*

rufipogon to the primarily inbreeding habit of *O. sativa* (Miyata et al., 2007) and altered photoperiod sensitivity has greatly expanded the geographic limits of rice cultivation (Yano et al., 2000; Kojima et al., 2002; Tadege et al., 2003; Doi et al., 2004).

As rice became an important food source, selection favored features of the grain that made it more nutritious for human consumption. This included often subtle, yet important changes in seed length (Fan et al., 2006; Takano-Kai et al., 2009), width (Shomura et al., 2008), and weight (Song et al., 2007; Wang et al., 2008). Larger seeds generally meant fewer propagules per plant, but provided a greater source of food for emerging sedentary societies. Selection for plants that produced the most grain led to a shift in the number of grains produced by a single plant, increasing the grain yields possible in a cropping season (Ashikari et al., 2005). Modern *O. sativa* exhibits an almost complete loss of pigmentation from the seed coverings relative to its ancestor, a change that may have had a pleiotropic effect on seed dormancy and resistance to insects and diseases, and may also have reduced the cooking time of the cultivar (Sweeney et al., 2006; Furukawa et al., 2007).

Finally, modern rice varieties have been altered in a variety of ways that made the grain more palatable to humans, including changes in the taste, texture, and fragrance of the grain. Since rice is consumed almost exclusively as a whole grain, the physical attributes of the grain are the chief determinants of quality to the consumer (Unnevehr et al., 1992; Fitzgerald et al., 2009). As the regions of the world that rely on rice as a staple food become more affluent, they demand higher quality rice (Juliano and Villareal, 1993), making improving grain quality the second most important objective in modern breeding programs (Juliano and Duff, 1991; Fitzgerald et al., 2009). These

grain quality modifications are often regionally specific, as they are intrinsically linked with the cultures that favor them (Fitzgerald et al., 2008). Two grain quality features for which the genetic determinants have been identified include glutinous endosperm (the "sticky" characteristic of sushi rice) (Wang et al., 1995), and the characteristic fragrance of high-quality Basmati and Jasmine varieties (Bradbury et al., 2005; Chen et al., 2008). While these grain quality changes were not likely selected during the transition from *O. rufipogon* to *O. sativa*, they remain good examples of how this important crop species continues to evolve according to human preferences.

Using the evolutionary history of rice to improve breeding efforts

Researchers have intensely studied the evolutionary history of rice since long before the genetic basis of inheritance was understood. This fascination with how modern rice evolved is not simply a retrospective look at the past, but promises to improve the ability of rice breeders to make selection decisions in the future. Understanding the genetic basis of the domestication syndrome has provided rice geneticists (and indeed all crop researchers) with a detailed look at what genetic factors are responsible for the morphological, physiological, and adaptive changes that made rice what it is today. Therefore, probing 10,000 years of rice evolution will provide useful information as we attempt to meet the breeding challenges we will be faced with in the future.

This dissertation attempts to further our understanding of the evolutionary history of rice by exploring specific genes related to rice domestication and varietal differentiation. It will first outline, based on existing evidence from multiple studies around the world, how genetic variation in *O. sativa* and *O. rufipogon* could best be exploited to achieve genetic gains (Chapter 2). Then, a deep synthesis of recent data on both genome-wide variation in rice and on cloned domestication-related genes will

be used to substantiate evolutionary patterns that are only now becoming evident (Chapter 3). An original research study exploring the origin of fragrance in rice will be presented to demonstrate the power of haplotype analysis for evolutionary studies and to provide new insights into varietal differentiation in rice (Chapter 4). The knowledge and resources from this study were also used to explore the origin of specific fragrance alleles in unique germplasm from Africa (Chapter 5) and Southeast Asia (Chapter 6). While many of the genes underlying domestication traits in rice have been identified, a major domestication trait remained to be characterized at the gene level: black hull coloration. Chapter 7 will present original research that mapped the genetic determinants of black hull, identified the causal mutation responsible for light hull in *O. sativa*, and confirmed that an epistatic interaction between two linked genes is responsible for this phenotype. Chapter 8 presents a more in-depth look at the sequence data from Chapter 4 in order to postulate the genetic and geographic origin of the major fragrance allele in rice. Finally, Chapter 9 will present a literature review and preliminary research that was done toward the development of a novel transgene containment strategy for rice.

REFERENCES

Ashikari M., Sakakibara H., Lin S., Yamamoto T., Takashi T., Nishimura A., Angeles E.R., Qian Q., Kitano H., Matsuoka M. (2005) Cytokinin oxidase regulates rice grain production. Science 309:741-745

Bradbury L.M.T., Fitzgerald T.L., Henry R.J., Jin Q., Waters D.L.E. (2005) The gene for fragrance in rice. Plant Biotechnology Journal 3:363-370

Cai, Cai H., Morishima, Morishima H. (2002) QTL clusters reflect character associations in wild and cultivated rice. Theoretical and Applied Genetics 104:1217-1228

Chen S., Yang Y., Shi W., Ji Q., He F., Zhang Z., Cheng Z., Liu X., Xu M. (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. Plant Cell 20:1850-1861

Doi K., Izawa T., Fuse T., Yamanouchi U., Kubo T., Shimatani Z., Yano M., Yoshimura A. (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT*-like gene expression independently of *Hd1*. Genes & Development 18:926-936

Fan C., Xing Y., Mao H., Lu T., Han B., Xu C., Li X., Zhang Q. (2006) *GS3* , a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics 112:1164-1171

Fitzgerald M.A., McCouch S.R., Hall R.D. (2009) Not just a grain of rice: the quest for quality. Trends in Plant Science 14:133-139

Fitzgerald M.A., Sackville Hamilton N.R., Calingacion M.N., Verhoeven H.A., Butardo V.M. (2008) Is there a second fragrance gene in rice? Plant Biotechnology Journal 6:416-423

Fuller D., Harvey E., Ling Q. (2007) Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. Antiquity 81:316-331

Furukawa T., Maekawa M., Oki T., Suda I., Iida S., Shimada H., Takamure I., Kadowaki K.-i. (2007) The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. The Plant Journal 49:91-102

Gepts P. (2004) Crop domestication as a long-term selection experiment. Plant Breeding Reviews 24:1-44

Gu X.-Y., Kianian S.F., Foley M.E. (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). Genetics 166:1503-1516

Gu X.-Y., Turnipseed E.B., Foley M.E. (2008) The *qSD12* locus controls offspring tissue-imposed seed dormancy in rice. Genetics 179:2263-2273

Gu X.Y., Kianian S.F., Foley M.E. (2006) Isolation of three dormancy QTLs as Mendelian factors in rice. Heredity 96:93-99

Hu F.Y., Tao D.Y., Sacks E., Fu B.Y., Xu P., Li J., Yang Y., McNally K., Khush G.S., Paterson A.H., Li Z.-K. (2003) Convergent evolution of perenniality in rice and sorghum. Proceedings of the National Academy of Sciences of the United States of America 100:4050-4054

Jin J., Huang W., Gao J.-P., Yang J., Shi M., Zhu M.-Z., Luo D., Lin H.-X. (2008) Genetic control of rice plant architecture under domestication. Nature Genetics 40:1365-1369

Juliano B.O., Duff B. (1991) Rice grain quality as an emerging research priority in national programs. Rice Grain Marketing and Quality Issues. International Rice Research Institute, pp 55-64

Juliano B.O., Villareal C.P. (1993) Grain quality evaluation of world rices. In: Singh R.K., Singh U.S., Khush G.S. (eds) Aromatic Rices. Science Publishers, Inc., Enfield, NH

Khush G.S. (1997) Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25-34

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant and Cell Physiology 43:1096-1105

Komatsu K., Maekawa M., Ujiie S., Satake Y., Furutani I., Okamoto H., Shimamoto K., Kyozuka J. (2003) *LAX* and *SPA*: Major regulators of shoot branching in rice. Proceedings of the National Academy of Sciences of the United States of America 100:11765-11770

Konishi S., Izawa T., Lin S.Y., Ebana K., Fukuta Y., Sasaki T., Yano M. (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392-1396

Li C., Zhou A., Sang T. (2006) Rice domestication by reducing shattering. Science 311:1936-1939

Li P., Zeng D., Liu X., Xu D., Gu D., Li J., Qian Q. (2003a) Mapping and characterization of a tiller-spreading mutant *lazy-2* in rice. Chinese Science Bulletin 48:2715-2717

Li X., Qian Q., Fu Z., Wang Y., Xiong G., Zeng D., Wang X., Liu X., Teng S., Hiroshi F., Yuan M., Luo D., Han B., Li J. (2003b) Control of tillering in rice. Nature 422:618-621

McCouch S.R., Kovach M.J., Sweeney M.T., Semon M. (2010) Gene flow and genetic isolation in crop evolution. In: Gepts P. (ed) Proceedings of the Harlan II

Symposium: Biodiversity in Agriculture: Domestication, Evolution, & Sustainability. Cambridge University Press, UC Davis, CA

Miyata M., Yamamoto T., Komori T., Nitta N. (2007) Marker-assisted selection and evaluation of the QTL for stigma exsertion under *japonica* rice genetic background. Theoretical and Applied Genetics 114:539-548

Shomura A., Izawa T., Ebana K., Ebitani T., Kanegae H., Konishi S., Yano M. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40:1023-1028

Song X.-J., Huang W., Shi M., Zhu M.-Z., Lin H.-X. (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nature Genetics 39:623-630

Sweeney M.T., Thomson M.J., Pfeil B.E., McCouch S. (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294

Tadege M., Sheldon C.C., Helliwell C.A., Upadhyaya N.M., Dennis E.S., Peacock W.J. (2003) Reciprocal control of flowering time by *OsSOC1* in transgenic Arabidopsis and by *FLC* in transgenic rice. Plant Biotechnology Journal 1:361-369

Takano-Kai N., Jiang H., Kubo T., Sweeney M., Matsumoto T., Kanamori H., Padhukasahasram B., Bustamate C., Yoshimura A., Doi K., McCouch S.R. (2009) Evolutionary history of GS3, a gene conferrign grain length in rice. Genetics In Press

Tan L., Li X., Liu F., Sun X., Li C., Zhu Z., Fu Y., Cai H., Wang X., Xie D., Sun C. (2008) Control of a key transition from prostrate to erect growth in rice domestication. Nature Genetics 40:1360-1364

Thomson M.J., Tai T.H., McClung A.M., Lai X.H., Hinga M.E., Lobos K.B., Xu Y., Martinez C.P., McCouch S.R. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theoretical and Applied Genetics 107:479-493

Unnevehr L.J., Duff B., Juliano B.O. (1992) Consumer demand for rice grain quality. Terminal Report IDRC Projects. National Grain Quality (Asia) and International Grain Quality Economics (Asia). International Rice Research Institute; International Development Research Centre, Manila, Philippines

Wang E., Wang J., Zhu X., Hao W., Wang L., Li Q., Zhang L., He W., Lu B., Lin H., Ma H., Zhang G., He Z. (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40:1370-1374

Wang Z.-Y., Zheng F.-Q., Shen G.-Z., Gao J.-P., Snustad D.P., Li M.-G., Zhang J.-L., Hong M.-M. (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. The Plant Journal 7:613-622

Xiong L.Z., Liu K.D., Dai X.K., Xu C.G., Zhang Q. (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theoretical and Applied Genetics 98:243-251

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., Sasaki T. (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 12:2473-2484

Yu Y., Tang T., Qian Q., Wang Y., Yan M., Zeng D., Han B., Wu C.-I., Shi S., Li J. (2008) Independent losses of function in a polyphenol oxidase in rice: Differentiation in grain discoloration between subspecies and the role of positive selection under domestication. Plant Cell 20:2946-2959

CHAPTER 2:
LEVERAGING NATURAL DIVERSITY:
BACK THROUGH THE BOTTLENECK

ABSTRACT

Plant breeders have long recognized the existence of useful genetic variation in the wild ancestors of our domesticated crop species. In cultivated rice (*Oryza sativa*), crosses between high-yielding elite cultivars and low-yielding wild accessions often give rise to superior offspring, with wild alleles conferring increased performance in the context of the elite cultivar genetic background. Because the breeding value of wild germplasm cannot be determined by examining the performance of wild accessions, a phylogenetic approach is recommended to determine which interspecific combinations are most likely to be useful in a breeding program. As we deepen our understanding of how genetic diversity is partitioned within and between cultivated and wild gene pools of *Oryza*, breeders will have increased power to make predictions about the most efficient strategies for utilizing wild germplasm for rice improvement.

INTRODUCTION

Approximately 10,000 years ago, Neolithic hunter–gatherers throughout Asia began to collect wild rice and impose unconscious selection, marking the beginning of a complex history of rice domestication. As people slowly tamed wild rice and learned to mold their societies around the requirements of rice production, they created the world's most enduring monoculture, and in turn became entirely dependent on this cereal for their daily sustenance. The global dependency on cultivated rice (*Oryza sativa*) has continued to strengthen, as now nearly half the world's population relies on rice as a staple food (IRGSP, 2005). But how did the unruly wild rice encountered by early human societies become transformed into the domesticated, high-yielding

varieties of rice we have today? This change was made possible by the existence of natural genetic variation, which humans harnessed by selecting for favorable traits. Modern breeding practices continue to follow in the footsteps of our ancestors as we seek new sources of genetic variation for rice improvement. This review will outline how breeders are making use of the natural diversity in *O. sativa* and discuss the usefulness of wild rice germplasm for rice improvement. A framework for generating the maximum amount of useful genetic novelty in an efficient and predictive manner will be presented, along with the challenges associated with this strategy.

Genetic bottlenecks and the evolution of population substructure in *O. sativa*

During the initial process of rice domestication, key traits such as diminished grain shattering and less persistent grain dormancy were strongly selected for by humans (Chang, 1995). Preferential propagation of individuals possessing these valuable traits would have created a primary domestication bottleneck. During this genetic bottleneck, many undesirable alleles from the wild ancestor, along with some potentially beneficial alleles, were not carried through to early domesticates (landraces), resulting in a narrowing of the domesticated rice gene pool (Tanksley and McCouch, 1997). Modern plant breeding continues to constrain the genetic diversity of cultivated rice by selecting for optimal performance under a highly managed set of agricultural conditions (Ladizinsky, 1985; McCouch, 2004). Therefore, modern cultivated rice is estimated to retain only approximately 10–20% of the genetic diversity present in its wild rice ancestor, *O. rufipogon* (Caicedo et al., 2007; Zhu et al., 2007).

Two genetically distinct groups within *O. sativa*, *Indica* and *Japonica*, have been recognized since ancient times [reviewed in (Chou, 1948)]. These two varietal groups

(sometimes referred to as subspecies) are believed to have been domesticated from geographically overlapping, yet genetically divergent populations of *O. rufipogon* (Figure 2.1), a concept that is extensively supported by molecular evidence (Ting, 1957; Oka, 1988; Chen et al., 1993; Lu et al., 2002; Cheng et al., 2003; Ma and Bennetzen, 2004; Hu et al., 2006; Londo et al., 2006; Caicedo et al., 2007; Rakshit et al., 2007). Further, sequence comparisons of chloroplast, mitochondrial and nuclear genomes suggest that the divergence of the *Indica* and *Japonica* gene pools predates the earliest archaeological evidence for rice domestication by 50–100,000 years (Ma and Bennetzen, 2004; Vitte et al., 2004; Zhu and Ge, 2005; Tang et al., 2006).

The two major varietal groups in *O. sativa* are further differentiated into five distinct subpopulations that can be clearly diagnosed using isozyme, simple sequence repeat (SSR), chloroplast, and/or single nucleotide polymorphism (SNP) markers (Glaszmann, 1987; Garris et al., 2005; Caicedo et al., 2007). Based on these studies, the *Japonica* varietal group can be divided into the *temperate japonica*, *tropical japonica*, and *Group V* (formerly *aromatic*) subpopulations while the *Indica* varietal group contains the *indica* and *aus* subpopulations. These subpopulations are well differentiated from each other, as indicated by pairwise F_{ST} values ranging from 0.2 to 0.42 (Garris et al., 2005), and they also differ in effective population size as a result of the proximity, duration and severity of the population bottlenecks experienced by each (Garris et al., 2005; Caicedo et al., 2007) (Figure 2.2).

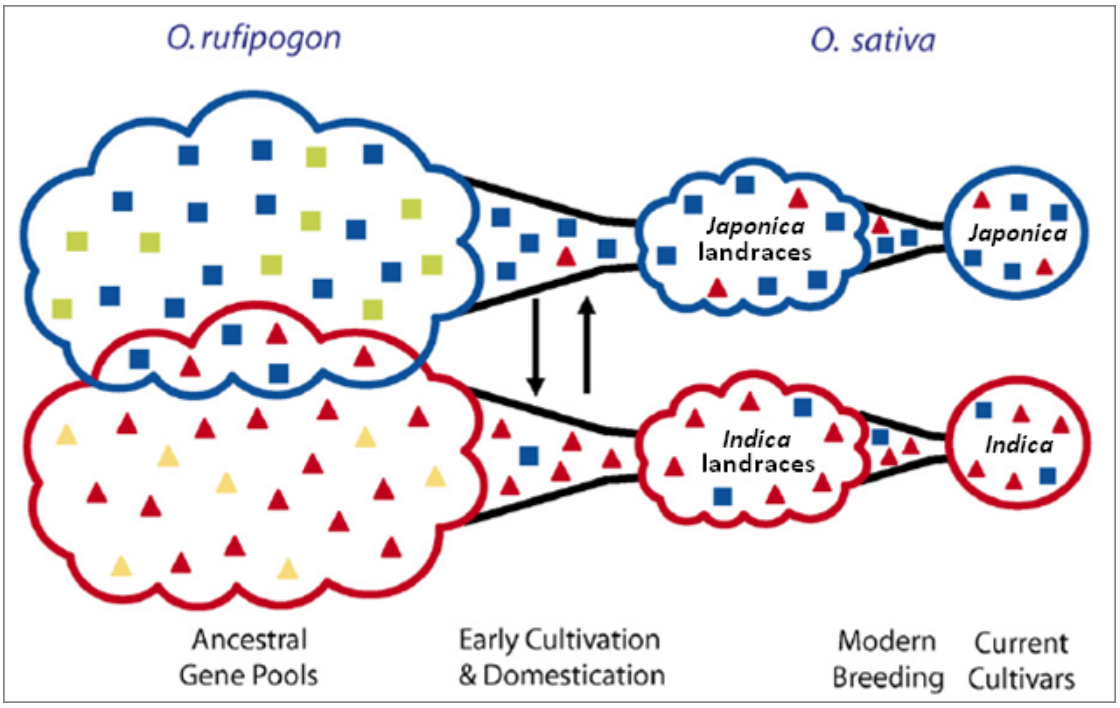


Figure 2.1: The complex domestication process of *O. sativa.* In contrast to the linear domestication bottleneck model proposed by Tanksley and McCouch (1997), the domestication process in *O. sativa* was considerably more complex. Phylogenetic, molecular, and archaeological evidence support the concept that diverse ancestral *O. rufipogon* populations existed over a broad geographical range across Asia and that multiple *O. rufipogon* populations gave rise to at least two primary domesticated varietal groups, namely the *Indica* and *Japonica* gene pools of domesticated rice. The cloud shapes represent the ancient gene pools of *O. rufipogon* and *O. sativa* that gave rise to modern cultivars. The triangle (*Indica*-specific) and square (*Japonica*-specific) shapes represent alleles; some of which (red/blue colors) were carried through the domestication bottleneck to modern cultivars, while others (yellow/green colors) were left behind in the wild species. Gene flow between early *Indica* and *Japonica* domesticates is depicted by arrows between the gene pools.

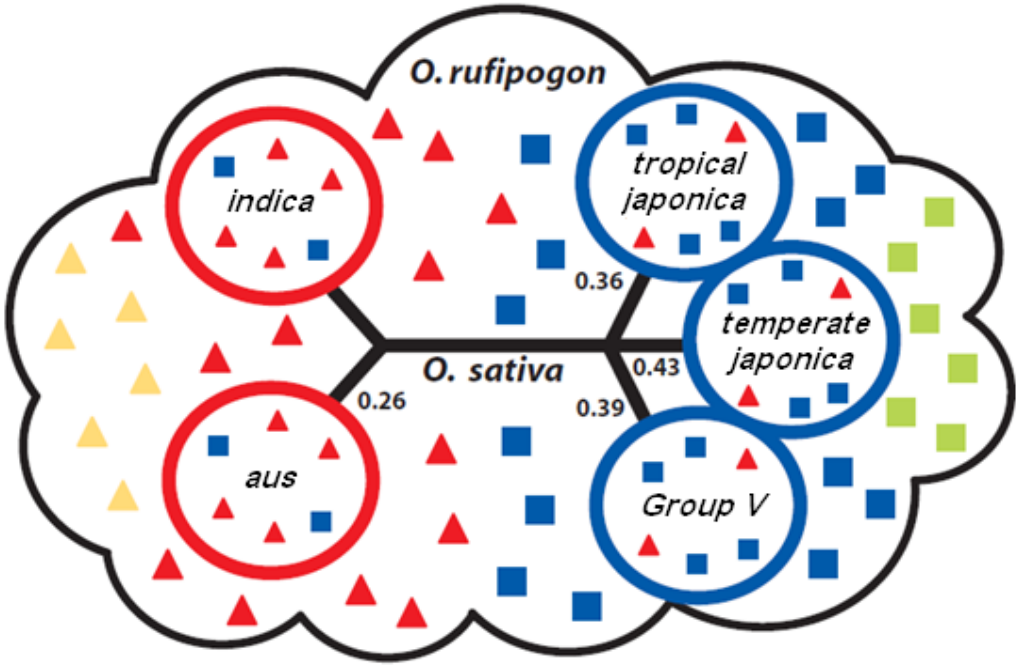


Figure 2.2: Subpopulation structure of *O. sativa*. In-depth genetic analysis has revealed that the two varietal groups in *O. sativa* (*Indica* and *Japonica*) are further subdivided into five distinct subpopulations: *indica* and *aus* (in the *Indica* varietal group) and *temperate japonica*, *tropical japonica*, and *Group V* (in the *Japonica* varietal group) (Garris et al., 2005; Caicedo et al., 2007). Circles representing the five *O. sativa* subpopulations are colored to indicate their relationship to the two varietal groups (*Indica* = red, *Japonica* = blue); domesticated sub-populations are superimposed over the large and diverse *O. rufipogon* ancestral gene pool where *Indica*- alleles are represented by triangles and *Japonica*-alleles are represented by squares; allele distribution within the *O. rufipogon* cloud indicates that some wild genotypes are more closely related to certain *O. sativa* genotypes than to each other. The five groups are highly differentiated from each other, as evidenced by the large F_{ST} values (Garris et al., 2005). Pairwise F_{ST} values relative to the *indica* subpopulation are indicated along the branches of the tree.

The process of rice domestication was accompanied by a gradual shift from the primarily out-crossing wild ancestor, *O. rufipogon*, to the primarily inbreeding domesticated species, *O. sativa*. This shift in mating system is largely responsible for the partitioning of genetic variation into the present-day subpopulations of *O. sativa*. Despite this shift in mating system, there is evidence of gene flow among the early *O. sativa* domesticates and between them and their sympatric wild relatives, infusing alleles that contributed in different ways to the diversity of each of the cultivated subpopulations (Oka, 1988). Several in-depth studies demonstrate that key domestication alleles are shared between the *Indica* and *Japonica* varietal groups (Yamanaka et al., 2004; Li et al., 2006; Sweeney et al., 2007), suggesting that rice domestication involved multiple genetic bottlenecks, coupled with episodes of hybridization and introgression between early rice domesticates from divergent gene pools (Kovach et al., 2007; Sang and Ge, 2007) (Figure 2.1).

Implications of subpopulation structure on rice breeding

The deep genetic differentiation among *O. sativa* subpopulations has several major implications for rice breeders. First, it is accompanied by intraspecific sterility barriers and reproductive incompatibilities that make it difficult to recover a full array of viable recombinant offspring when crosses are made between the *Indica* and *Japonica* varietal groups (Oka, 1988; Harushima et al., 2002). As a result, rice breeders have historically focused on crosses between genotypes within a varietal group (i.e. *temperate japonica* x *tropical japonica*; *indica* x *indica*) (Sano, 1993; Ni et al., 2002; Lu et al., 2005).

A more provocative implication of the subpopulation structure in rice is that it provides rice breeders with a suite of naturally occurring, highly divergent gene pools

that can appropriately be considered 'heterotic' or 'combinability' groups. In the traditional sense, heterotic groups are populations of a species that are sufficiently divergent so that when crosses are made between the groups, a significant amount of heterosis (hybrid vigor) may be observed in the F_1 generation (Allard, 1960). In rice, several decades of work in China have demonstrated that F_1 hybrids derived from crosses between divergent rice subpopulations are generally more productive than F_1 hybrids derived from crosses between closely related cultivars, as long as the sterility barriers are carefully managed (Li and Yuan, 2000; Cheng et al., 2007).

Divergent rice populations have also been proposed to function as combinability groups for exploiting transgressive variation during the development of superior inbred varieties (McCouch et al., 2007). Transgressive variation is a phenomenon that is recognized by the appearance of individuals in the progeny of a cross that exceed the performance of the better parent. It is generally explained by the fact that most lines contain a distribution of both positive and negative alleles that contribute to an intermediate phenotype. When genetically divergent parents are crossed, recombination in the offspring gives rise to segregants (individual progeny) that are more extreme than either parent because they contain higher frequencies of either favorable or unfavorable alleles. The exploitation of this phenomenon has guided the improvement of inbred varieties where divergent germplasm resources are used to generate transgressive variation, which breeders can then fix in the elite backgrounds of interest (Li et al., 2005; Tian et al., 2006a; McCouch et al., 2007).

Wild germplasm as a resource for capturing positive transgressive variation

Rice breeders today face the formidable challenge of achieving the pest resistance, stress tolerance, yield, and quality improvements that will be necessary to keep pace

with rising global food requirements. The probability of success in this endeavor depends to a great extent on our ability to make use of novel sources of genetic variation. One way to do this is to explore the largely untapped reservoir of allelic diversity that remains hidden within existing populations of early landraces and wild relatives. New technology makes it possible to readily identify wild alleles that were left behind by ancient farmers and to selectively harness those that enhance performance when introduced into our highly productive modern varieties.

Landraces of *O. sativa* are genetic intermediates between wild ancestors and modern, elite cultivars. Having been selected for alleles and adaptive gene complexes that are favorable to humans, they represent a rich pool of genetic diversity that is readily accessible to modern rice breeders. Why then should we look to the poor performing, low-yielding rice ancestors for the novelty necessary for rice improvement? One reason is that all elite cultivars are the result of selections from landrace varieties that have themselves been through the primary domestication bottleneck. Modern varieties are therefore expected to share a higher proportion of alleles with landraces than with wild accessions. Thus, the probability of generating novel genetic variation from crosses between elite varieties and wild germplasm is greater than would be expected from crosses to landrace materials. In addition, crosses between elite cultivars and wild germplasm generally present fewer reproductive barriers than do crosses between *Indica* and *Japonica* cultivars (Oka, 1988; Harushima et al., 2002; McCouch et al., 2007).

Numerous studies report improvements in performance because of the introgression of valuable genes from wild germplasm into elite rice cultivars. Historically, breeders identified phenotypes such as disease resistance or male sterility in a wild rice species

and then introduced the trait through backcross breeding [reviewed in (Brar and Khush, 1997)]. More recently, the use of advanced backcross quantitative trait locus (QTL) analysis and near isogenic lines (NILs) have made it technically and economically feasible to identify and selectively introgress genes or QTL that confer superior performance in the genetic background of an elite cultivar but that have no observable phenotype in the wild donor (Figure 2.3). Despite its inferior yield and agronomic performance, *O. rufipogon* has been the source of beneficial alleles for diverse quantitative traits including grain size, grain weight (Li et al., 2004a; Xie et al., 2006), grain yield (Xiao et al., 1996; Xiao et al., 1998; Moncada et al., 2001; Li et al., 2002; Septiningsih et al., 2003a; Thomson et al., 2003; Marri et al., 2005; He et al., 2006; Tian et al., 2006a; Tian et al., 2006b; Xie et al., 2006; McCouch et al., 2007), grain quality (Septiningsih et al., 2003b), cold tolerance (Liu et al., 2003), aluminum tolerance (Nguyen et al., 2003), and flowering time (Thomson et al., 2006). Yield and grain quality enhancing alleles have also been identified from *O. glaberrima* (Aluko et al., 2004; Li et al., 2004b; Sarla and Swamy, 2005) and *O. glumaepatula* (Brondani et al., 2002; Rangel et al., 2005).

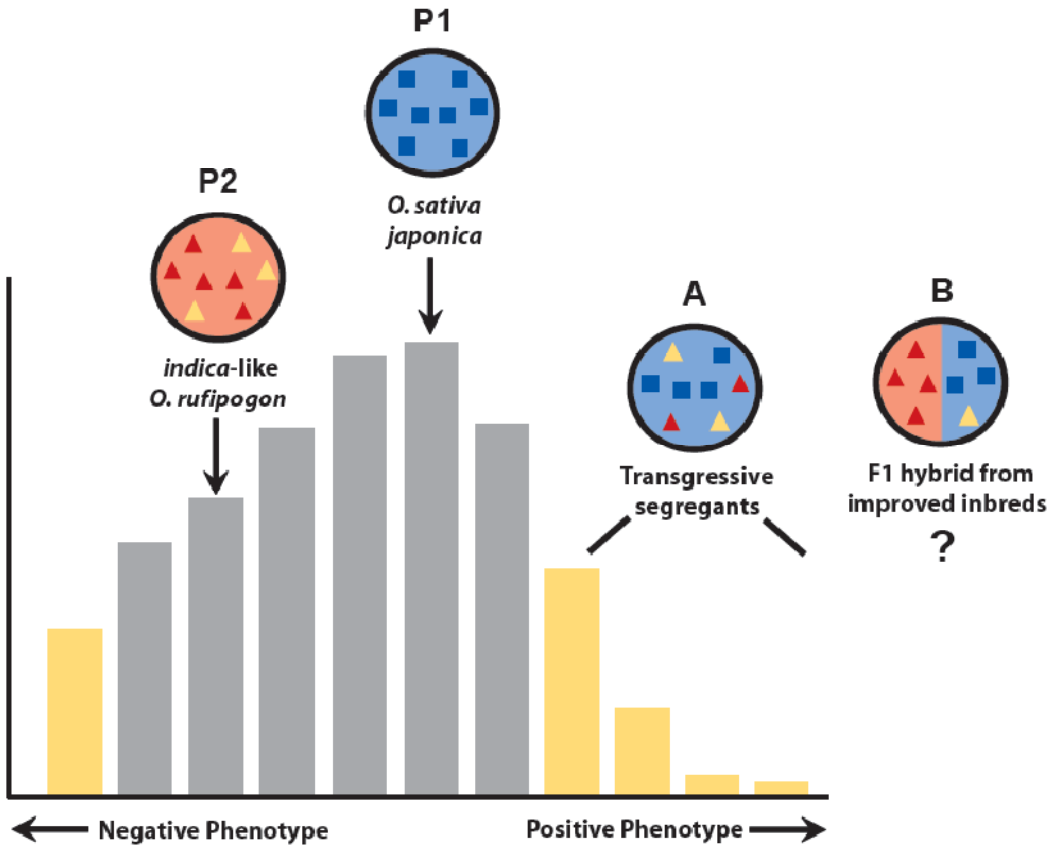


Figure 2.3: Transgressive Segregation. Transgressive segregation is observed in crosses between genetically divergent genotypes where the progeny exceed the performance of the parents. This Figure displays the phenotypic distribution of a hypothetical cross between an elite *Japonica* cultivar (P1) and an *Indica*-like *O. rufipogon* accession (P2). Since the two parents do not share many of the same alleles, there is a high probability that some of the progeny (A) will possess novel combinations of alleles that confer a performance advantage. Selective introgression of beneficial alleles from exotic germplasm sources offers a way to expand the gene pool of modern cultivars without disrupting many of the gene complexes that contribute to the quality and adaptation of elite cultivars. Divergent introgression lines could then be used as parents to create F1 hybrids (B) that aim to maximize the heterotic potential of *O. sativa*.

In cases where genes are introgressed from genetically divergent, low-performing wild or weedy donors, the alleles of interest are associated with positive transgressive variation in elite genetic backgrounds (Figure 2.4). This phenomenon has been demonstrated through interspecific crosses in many crop species (Rick, 1974; Cox and Frey, 1984; deVicente and Tanksley, 1993; Eshed and Zamir, 1995; Rieseberg et al., 1999; Rao et al., 2003; Lacape et al., 2005), highlighting the potential usefulness of exploring exotic germplasm sources for the improvement of a wide array of domesticated crop species.

Making the most of transgressive variation in rice

One method of delivering superior varieties to farmers is through F_1 hybrid technology. In China, the superior performance of hybrid versus inbred rice varieties has resulted in the expansion of hybrid production to approximately 50% of the total rice production (Cheng et al., 2007). For the last 80 years, hybrid technology has been largely responsible for the steady increase in maize yields in the United States (1% per year) and more recently for sorghum as well (Duvick, 1999; Duvick and Cassman, 1999; Duvick, 2005). In light of these successes, and of the industry's interest in the hybrid model, it is tempting to assume that hybrids will drive increases in productivity for most crops in the future. However, the relative costs and benefits of hybrid versus inbred variety development in inbreeding versus out-crossing species, and in high value versus low-value crops, suggest that the answer is not entirely clear.

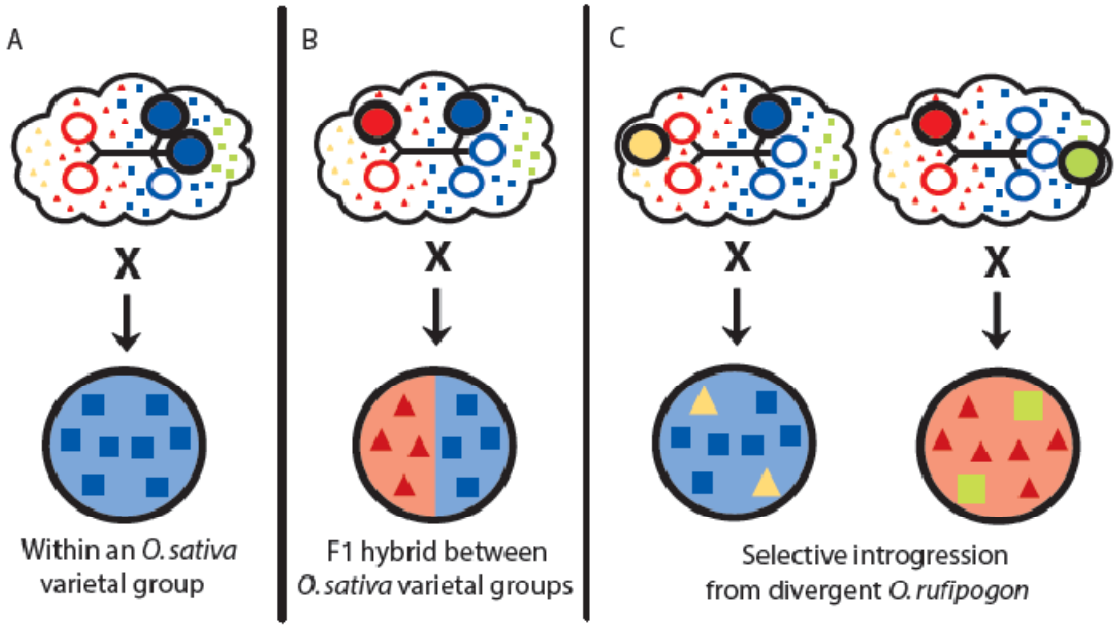


Figure 2.4: Rice breeding options: How to generate novelty? Traditional rice breeding has generated elite cultivars derived from crosses between genetically similar germplasm, such as between members of the same varietal group (panel A). While this avoids potential problems with reproductive barriers and quality issues, little genetic novelty is available for enhancing the performance of the cultivar. In contrast, F1 hybrids between genetically divergent groups, such as between *Indica* and *Japonica* parents, bring together alleles that contribute to heterosis (panel B). A third option is to selectively introgress genes from genetically divergent germplasm (i.e., from *O. rufipogon*) into elite *O. sativa* cultivars, creating introgression lines that exhibit positive transgressive variation (panel C). It is of interest to determine whether carefully crafted introgression lines make it possible for inbred varieties to equal or outperform F1 hybrids.

The small filled circles with a bold outline at the top of the panel indicate the germplasm pools being crossed (corresponding to the subpopulations indicated in Figure 2.2). The large circles at the bottom of each panel represent individual genotypes resulting from a particular cross; the red triangles and blue squares represent *Indica*-like and *Japonica*-like alleles, respectively; yellow triangles and green squares represent alleles from *O. rufipogon* that were left behind during the domestication bottleneck, but are re-introduced into inbred elite lines through selective backcrossing. Panel C shows two introgression lines, where divergent *O. rufipogon* alleles were introduced into a *Japonica* (blue) and an *Indica* (red) genetic background.

The perfect, cleistogamous flowers of rice make it difficult and costly to reliably obtain out-crossed F_1 seed, even with the development of both two- and three-line male sterility systems (Virmani and Kumar, 2004). In addition, genetic evidence suggests that overdominance does not appear to be the major cause of heterosis in *O. sativa*, making it possible to capture complementary alleles in inbred varieties (Xiao et al., 1995; Yu et al., 1997; Hua et al., 2002; Hua et al., 2003). Further, because inbreeding crops have expunged most deleterious recessive alleles over the course of evolution, the heterozygosity provided by F_1 hybrids is not theoretically necessary to achieve superior performance (Xiao et al., 1995; Hua et al., 2002).

Thus, we propose that rice breeders can capture a large portion of heterosis in inbred varieties and that wild germplasm represents an underutilized source of novel alleles. To approach this systematically, genetic diversity within *O. rufipogon* must be characterized and the relationships between subpopulations of *O. sativa* and *O. rufipogon* must be defined. Several studies have reported ecological and/or geographical population substructure in *O. rufipogon*. While the relationship between the wild and cultivated subpopulations is complicated by the substantial and well-documented gene flow between them (Sun et al., 2002; Ishikawa et al., 2006), there are subpopulations of *O. rufipogon* that cluster nearer to some subpopulations of *O. sativa* than others (Lu et al., 2002; Cheng et al., 2003; Hu et al., 2006; Londo et al., 2006; Rakshit et al., 2007) (Figure 2.2). We can, therefore, use a phylogenetic approach to select wild genotypes that are genetically divergent from target elite cultivars to use as parents. This strategy will help to maximize the probability of creating useful transgressive segregation from which to select superior phenotypes (Figure 2.3). By introgressing a few, selected chromosomal segments (QTLs) from genetically divergent wild donors, breeders aim to move existing elite rice cultivars

'up the fitness landscape' (Cooper et al., 2005), fixing positive transgressive segregants through repeated backcrossing and selfing (McCouch et al., 2007). Once these 'wild QTLs' are fixed in improved inbred varieties, they may also be useful to hybrid breeders who can take advantage of them to create a new generation of superior hybrids (Ma and Yuan, 2004) (Figure 2.4).

Conclusions—future challenges

The immediate wild ancestor of rice, *O. rufipogon*, is known to contain alleles that confer valuable transgressive variation when introgressed into elite cultivars of *O. sativa*, but at this time there is no predictive model that will tell us, *a priori*, where to look for the valuable wild alleles. The challenge before us is therefore to integrate information from both whole genome SNP assays and targeted gene-mapping studies as a step toward more efficient utilization of wild relatives for rice improvement. New technologies can now be used to re-sequence entire genomes and to define regions that are highly divergent between gene pools or regions that are shared. This will allow us to identify genomic segments that are common by descent in both *Indica* and *Japonica* but divergent in *O. rufipogon* (and may correspond to domestication loci), as well as regions of admixture between populations. Chromosomal regions of interest can be introgressed into a suite of elite cultivars to determine whether estimates of divergence are predictive of positive transgressive variation following hybridization. As we gain knowledge about the genes, functional nucleotide polymorphisms and pathways underlying positive transgressive variation, we will be able to make and test predictions about how specific genes or alleles will interact with each other in a given genetic background. We will also be able to examine the relationship between SNP frequencies, genome wide patterns of diversity and patterns of linkage disequilibrium to make and test predictions about which of the many wild or exotic accessions

combine best with specific elite materials. This information will lay the foundation for 'reverse genetics' models that allow us to more efficiently utilize the wealth of natural variation that resides on the other side of the domestication bottleneck.

REFERENCES

Allard R.W. (1960) Principles of Plant Breeding. John Wiley and Sons, Inc., New York

Aluko G., Martinez C., Tohme J., Castano C., Bergman C., Oard J.H. (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theoretical and Applied Genetics 109:630-639

Brar D.S., Khush G.S. (1997) Alien introgression in rice. Plant Molecular Biology 35:35-47

Brondani, Brondani C., Rangel, Rangel P., Brondani R., Ferreira, Ferreira M. (2002) QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theoretical and Applied Genetics 104:1192-1203

Caicedo A.L., Williamson S.H., Hernandez R.D., Boyko A., Fledel-Alon A., York T.L., Polato N.R., Olsen K.M., Nielsen R., McCouch S.R., Bustamante C.D., Purugganan M.D. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genetics 3:e163

Chang T.T. (1995) Rice. In: Smartt J.S., N.W. (ed) Evolution of Crop Plants. Longman Scientific and Technical

Chen W.B., Nakamura I., Sato Y.I., Nakai H. (1993) Distribution of deletion type in cpDNA of cultivated and wild rice. Japanese Journal of Genetics 68:597-603

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67-75

Cheng S.-H., Zhuang J.-Y., Fan Y.-Y., Du J.-H., Cao L.-Y. (2007) Progress in research and development on hybrid rice: A super-domesticate in China. Annals of Botany 100:959-966

Chou S.L. (1948) China is the place of origin of rice. J Rice Soc 7:53-54

Cooper M., Podlich D.W., Smith O.S. (2005) Gene-to-phenotype models and complex trait genetics. Australian Journal of Agricultural Research 56:895-918

Cox D.J., Frey K.J. (1984) Improving cultivated oats (*Avena sativa* L.) with alleles for vegetative growth index from *A. sterilis* L. Theoretical and Applied Genetics 68:239-245

deVicente M.C., Tanksley S.D. (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585-596

Duvick D.N. (1999) Feeding people and protecting natural resources. In: Coors J.G.P., S. (ed) The Genetics and Exploitation of Heterosis in Crops. Crop Science Society of America and Soil Science Society of America, pp 19-29

Duvick D.N. (2005) The contribution of breeding to yield advances in maize (*Zea mays* L.). Advances in Agronomy 86:83-145

Duvick D.N., Cassman K.G. (1999) Post-Green Revolution trends in yield potential of temperate maize in the North-Central United States. Crop Science 39:1622-1630

Eshed Y., Zamir D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147-1162

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638

Glaszmann J.C. (1987) Isozymes and classification of Asian rice varieties. Theoretical and Applied Genetics 74:21-30

Harushima Y., Nakagahra M., Yano M., Sasaki T., Kurata N. (2002) Diverse variation of reproductive barriers in three intraspecific rice crosses. Genetics 160:313-322

He G., Luo X., Tian F., Li K., Zhu Z., Su W., Qian X., Fu Y., Wang X., Sun C., Yang J. (2006) Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. Genome Research 16:618-626

Hu H., Mu J., Zhang H.-J., Tao Y.-Z., Han B. (2006) Differentiation of a miniature inverted transposable element (MITE) system in Asian rice cultivars and its inference for a diphyletic origin of two subspecies of Asian cultivated rice. Journal of Integrative Plant Biology 48:260-267

Hua J.P., Xing Y.Z., Xu C.G., Sun X.L., Yu S.B., Zhang Q. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885-1895

Hua J., Xing Y., Wu W., Xu C., Sun X., Yu S., Zhang Q. (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proceedings of the National Academy of Sciences of the United States of America 100:2574-2579

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793-800

Ishikawa R., Yamanaka S., Fukuta Y., Chitrakon S., Bounphanousay C., Kanyavong K., Tang L., Nakamura I., Sato T., Sato Y.I. (2006) Genetic erosion from modern varieties into traditional upland rice cultivars. Genetic Resources and Crop Evolution 53:245-252

Kovach M.J., Sweeney M.T., McCouch S.R. (2007) New insights into the history of rice domestication. Trends in Genetics 23:578-587

Lacape J.-M., Nguyen T.-B., Courtois B., Belot J.-L., Giband M., Gourlot J.-P., Gawryziak G., Roques S., Hau B. (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* x *Gossypium barbadense* backcross generations. Crop Science 45:123-140

Ladizinsky G. (1985) Founder effect in crop-plant evolution. Economic Botany 39:191-199

Li C., Zhou A., Sang T. (2006) Rice domestication by reducing shattering. Science 311:1936-1939

Li D., Sun C., Fu Y., Li C., Zhu Z., Chen L., Cai H., Wang X. (2002) Identification and mapping of genes for improving yield from Chinese common wild rice (O. rufipogon Griff.) using advanced backcross QTL analysis. Chinese Science Bulletin 18:1533-1537

Li J., Thomson M., McCouch S.R. (2004a) Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168:2187-2195

Li J., Xiao J., Grandillo S., Jiang L., Wan Y., Deng Q., Yuan L., McCouch S.R. (2004b) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome 47:697-704

Li J., Yuan L. (2000) Hybrid rice: genetics, breeding and seed production. Plant Breeding Reviews 17:150-158

Li Z.-K., Fu B.-Y., Gao Y.-M., Xu J.-L., Ali J., Lafitte H., Jiang Y.-Z., Rey J., Vijayakumar C., Maghirang R., Zheng T.-Q., Zhu L.-H. (2005) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Molecular Biology 59:33-52

Liu F., Sun C., Tan L., Fu Y., Li D., Wang X. (2003) Identification and mapping of quantitative trait loci controlling cold-tolerance of Chinese common wild rice (*O. rufipogon* Griff.) at booting to flowering stages. Chinese Science Bulletin 48:2068-2071

Londo J.P., Chiang Y.-C., Hung K.-H., Chiang T.-Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proceedings of the National Academy of Sciences 103:9578-9583

Lu, Lu B.R., Zheng, Zheng K., Qian, Qian H., Zhuang, Zhuang J. (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics 106:101-106

Lu H., Redus M.A., Coburn J.R., Rutger J.N., McCouch S.R., Tai T.H. (2005) Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis. Crop Science 45:66-76

Ma G., Yuan L. (2004) Hybrid rice achievements and development in China. In: Virmani S.S.M., C.X.; Hardy, B. (ed) Proceedings of the 4th International Symposium on Hybrid Rice, Los Banos, Philippines International Rice Research Institute, pp 247-257

Ma J., Bennetzen J.L. (2004) Rapid recent growth and divergence of rice nuclear genomes. Proceedings of the National Academy of Sciences of the United States of America 101:12404-12410

Marri P., N S., Reddy L., Siddiq E. (2005) Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogon*. BMC Genetics 6:33

McCouch S. (2004) Diversifying selection in plant breeding. PLoS Biology 2:e347

McCouch S., Sweeney M., Li J., Jiang H., Thomson M., Septiningsih E., Edwards J., Moncada P., Xiao J., Garris A., Tai T., Martinez C., Tohme J., Sugiono M., McClung A., Yuan L., Ahn S.-N. (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica 154:317-339

Moncada P., Martínez C.P., Borrero J., Chatel M., Gauch Jr H., Guimaraes E., Tohme J., McCouch S.R. (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa*×*Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theoretical and Applied Genetics 102:41-52

Nguyen B., Brar D., Bui B., Nguyen T., Pham L., Nguyen H. (2003) Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into *indica* rice (*Oryza sativa* L.). Theoretical and Applied Genetics 106:583-593

Ni J., Colowit P.M., Mackill D.J. (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Science 42:601-607

Oka H.I. (1988) Origin of Cultivated Rice. Elsevier Science/Japan Scientific Societies Press, Tokyo

Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N., Terauchi R. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theoretical and Applied Genetics 114:731-743

Rangel P.H.N., Brondani C., Rangel P.N., Brondani R.P.V., Zimmermann F.J.P. (2005) Development of rice lines with gene introgression from the wild *Oryza glumaepatula* by the AB-QTL methodology. Crop Breed Appl Biotech 5:10-19

Rao G.U., Ben Chaim A., Borovsky Y., Paran I. (2003) Mapping of yield-related QTLs in pepper in an interspecific cross of *Capsicum annuum* and *C. frutescens*. Theoretical and Applied Genetics 106:1457-1466

Rick C.M. (1974) High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species. Hilgarida 42:493-510

Rieseberg L.H., Archer M.A., Wayne R.K. (1999) Transgressive segregation, adaptation and speciation. Heredity 83:363-372

Sang T., Ge S. (2007) The puzzle of rice domestication. Journal of Integrative Plant Biology 49:760-768

Sano Y. (1993) Constraints in using wild relatives in breeding: lack of basic knowledge on crop gene pools. In: Buxton D.R. (ed) International Crop Science. Crop Science Society of America, pp 437-443

Sarla N., Swamy B.P.M. (2005) *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. Current Science 89:955-963

Septiningsih E.M., Prasetiyono J., Lubis E., Tai T.H., Tjubaryat T., Moeljopawiro S., McCouch S.R. (2003a) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theoretical and Applied Genetics 107:1419-1432

Septiningsih E.M., Trijatmiko K.R., Moeljopawiro S., McCouch S.R. (2003b) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. Theoretical and Applied Genetics 107:1433-1441

Sun C., Wang X., Yoshimura A., Doi K. (2002) Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). Theoretical and Applied Genetics 104:1335-1345

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D., McCouch S.R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3:e133

Tang T., Lu J., Huang J., He J., McCouch S.R., Shen Y., Kai Z., Purugganan M.D., Shi S., Wu C.-I. (2006) Genomic variation in rice: Genesis of highly polymorphic linkage blocks during domestication. PLoS Genetics 2:e199

Tanksley S.D., McCouch S.R. (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066

Thomson M.J., Edwards J.D., Septiningsih E.M., Harrington S.E., McCouch S.R. (2006) Substitution mapping of *dth1.1*, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. Genetics 172:2501-2514

Thomson M.J., Tai T.H., McClung A.M., Lai X.H., Hinga M.E., Lobos K.B., Xu Y., Martinez C.P., McCouch S.R. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theoretical and Applied Genetics 107:479-493

Tian F., Li D., Fu Q., Zhu Z., Fu Y., Wang X., Sun C. (2006a) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theoretical and Applied Genetics 112:570-580

Tian F., Zhu Z., Zhang B., Tan L., Fu Y., Wang X., Sun C. (2006b) Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (*Oryza rufipogon* Griff.). Theoretical and Applied Genetics 113:619-629

Ting Y. (1957) The origin and evolution of cultivated rice in China. Acta Agron Sinica 8:243-260

Virmani S.S., Kumar I. (2004) Advances in hybrid rice research and development in the tropics. In: Virmani S.S.M., C.X.; Hardy, B. (ed) Proceedings of the 4th International Symposium on Hybrid Rice, Los Banos, Philippines International Rice Research Institute, pp 10-20

Vitte C., Ishii T., Lamy F., Brar D., Panaud O. (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Molecular Genetics and Genomics 272:504-511

Xiao J., Grandillo S., Ahn S.N., McCouch S.R., Tanksley S.D., Li J., Yuan L. (1996) Genes from wild rice improve yield. Nature 384:223-224

Xiao J., Li J., Grandillo S., Ahn S.N., Yuan L., Tanksley S.D., McCouch S.R. (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899-909

Xiao J., Li J., Yuan L., Tanksley S.D. (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745-754

Xie X., Song M.-H., Jin F., Ahn S.-N., Suh J.-P., Hwang H.-G., McCouch S. (2006) Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. Theoretical and Applied Genetics 113:885-894

Yamanaka S., Nakamura I., Watanabe K., Sato Y.-I. (2004) Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theoretical and Applied Genetics 108:1200-1204

Yu S.B., Li J.X., Xu C.G., Tan Y.F., Gao Y.J., Li X.H., Zhang Q., Maroof M.A.S. (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proceedings of the National Academy of Sciences of the United States of America 94:9226-9231

Zhu Q., Ge S. (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytologist 167:249-265

Zhu Q., Zheng X., Luo J., Gaut B.S., Ge S. (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. Molecular Biology and Evolution 24:875-888

CHAPTER 3:

NEW INSIGHTS INTO THE HISTORY OF RICE DOMESTICATION

ABSTRACT

The history of rice domestication has long been the subject of debate. Recently obtained genetic evidence provides us with new insights into this complex story. Genome-wide studies of variation demonstrate that the two varietal groups in *Oryza sativa, Indica* and *Japonica,* arose from genetically distinct gene pools within a common wild ancestor, *Oryza rufipogon*, suggesting multiple domestications of *O. sativa*. However, the evolutionary history of recently cloned domestication genes adds another layer of complexity to the domestication of rice. Although some alleles exist only within specific subpopulations, as would be expected if the domestications occurred independently, other major domestication alleles are common to all cultivated *O. sativa* varieties. Our current view of rice domestication supports multiple domestications coupled with limited introgression that transferred key domestication alleles between divergent rice gene pools.

From wild grass to staple crop

More than 10,000 years ago, ancient peoples began to gather and consume *Oryza rufipogon*, an unruly wild grass species that grew in the swamps and marshes throughout tropical and sub-tropical Asia (Box 3.1). Through a process of continuous selection for desirable features, these early farmers slowly transformed wild rice into *O. sativa*, which is now an essential staple crop for billions of people worldwide. Indeed, rice is now the primary source of food and livelihood for more than a third of

the world's population and is produced on every continent with arable land (Khush, 1997).

BOX 3.1: The wild ancestor of *O. sativa*

The wild ancestors of our cultivated crop species not only represent valuable sources of genetic diversity for crop improvement, but also provide a window through which we can catch a glimpse of the evolutionary history of our domesticated species. Several lines of evidence suggest that *O. sativa* was domesticated from *O. rufipogon*, or Asian common wild rice. Numerous genetic studies using molecular markers or DNA sequence information demonstrate that *O. sativa* cultivars are more closely related to *O. rufipogon* than to any other *Oryza* species, with *O. rufipogon* accessions often grouping together with *O. sativa* in phylogenetic analyses (Lu et al., 2002a; Rakshit et al., 2007). The geographic ranges of *O. rufipogon* and *O. sativa* overlap completely, and there are no major reproductive barriers between these two species, leading to a continuous series of intermediate admixed genotypes (Oka, 1988). These findings clearly demonstrate the close relationship between this wild species and cultivated rice. *O. rufipogon* currently thrives over a large geographic range from the Indian subcontinent through Southeast Asia and China and south to Oceania (Chang, 1995; Khush, 1997), although the range of this species in ancient times remains poorly defined.

O. rufipogon is a complex species that includes both annual (*O. nivara*) and perennial (*O. rufipogon*) forms. Some researchers recognize these as two separate species (Yamanaka et al., 2004; Li et al., 2006a), while others consider them to be ecotypes of a single species because they exhibit continuous variation in nature for their life history habit and no reproductive barrier exists between them (Oka, 1988; Morishima, 2001; Lu et al., 2002a). In addition, there is no significant genetic differentiation between *O. rufipogon* and *O. nivara*, as demonstrated by the abundance of shared polymorphisms, lack of fixed differences, and very low F_{st} values (Lu et al., 2002a; Ma and Bennetzen, 2004; Zhu et al., 2007). As such, we will treat these two species as a single large gene pool referred to as *O. rufipogon*.

Domestication involves a series of profound genetic changes resulting from selection on a wild species that make it more amenable for cultivation and consumption by humans. It is widely recognized that domestication is not a single "event", but rather a dynamic evolutionary process that occurs over time and, in some species, continues to this day (Gepts, 2004). The traits that distinguish modern rice varieties from their wild ancestor can range from subtle to dramatic (Figure 3.1). In addition to traits that

resulted in major alterations of plant structure and/or reproductive physiology, humans have selected for characteristics that made rice grains more appealing as a food source, including grain size, shape, color, fragrance and amylose content.

Another feature of domestication traits is that they are generally quantitative in nature, meaning that they are under the control of numerous genetic components that often interact in complex biochemical and regulatory pathways. For example, in rice there are at least five grain shattering loci (Kinoshita, 1998; Xiong et al., 1999; Gu et al., 2005b), a minimum of six dormancy loci (Gu et al., 2004, 2005a), as many as 31 loci associated with the change from outcrossing to inbreeding (Uga et al., 2003; Miyata et al., 2007), more than eight loci associated with grain weight (Thomson et al., 2003) and at least five loci associated with grain color (Kato and Ishikawa, 1921; Sweeney et al., 2006; Furukawa et al., 2007). Despite the quantitative nature of domestication traits, there appear to be a few key discrete genetic loci responsible for the major shift from wild to cultivated forms (Doebley et al., 2006). In addition to these major quantitative trait loci (QTLs), numerous genetic modifiers, or minor QTLs, contribute to the range of phenotypic variation for a given trait. To date (November 2007), all cloned domestication-related loci in rice have involved reproductive traits (Table 3.1).

Figure 3.1: The domestication transformation—From *O. rufipogon* to *O. sativa.* The wild ancestor of cultivated rice (*O. sativa*) is *O. rufipogon*, a diverse species that exists over broad geographic and ecological regions across Asia. During the domestication of rice, an entire suite of morphological and physiological traits were altered in response to human selection. Compared to its wild *O. rufipogon* ancestor, cultivated rice typically exhibits reduced grain shattering, reduced dormancy, loss of pigmentation in the hull and seed coat and a reduction in the rate of outcrossing. Modern rice varieties also display increased synchronization of tiller development and panicle formation, more secondary panicle branches, increased grain number and weight and a modified photoperiodic response (Chang, 1995). (a) Panicle from *O. rufipogon*; (b) seeds from *O. rufipogon*; (c) panicles from *O. sativa*; and (d) seeds from *O. sativa*.

Table 3.1: Key Domestication-Related Genes Cloned in Rice

Gene	Chr[a]	Trait Affected	Functional Mutation	Molecular Function	Ref.
Sh4	4	Grain shattering	SNP causing amino acid substitution	Myb3 transcriptional regulator	[1,2]
qSH1	1	Grain shattering	SNP in regulatory region	BEL1-type homeobox transcriptional regulator	[3]
Rc	7	Grain pericarp color	Deletion causing protein truncation	Basic helix-loop-helix transcriptional regulator	[4]
Waxy	6	"Sticky" (glutinous) grains	SNP in intron affecting mRNA splicing	Granule-bound starch synthase	[5-7]
GS3	3	Grain size/shape	SNP causing protein truncation	Cellular signaling protein with a VWFC module	[8]
[BADH2][b]	8	Grain fragrance/ flavor	Deletion causing protein truncation	Betaine aldehyde dehydrogenase	[9]
[Gn1a][b]	1	Grain number	Several possible mutations	Cytokinin oxidase/ dehydrogenase	[10]
[GW2][b]	2	Grain weight/width	Deletion causing protein truncation	RING-type protein with E3 ubiquitin ligase activity	[11]

[a]Chr. = chromosome

[b]The bracketed genes are genes that affect domestication-related traits, but there is no evidence at this time to determine whether these genes were part of the domestication process or whether they are novel mutations that have recently become the targets of selection

1- (Li et al., 2006b)
2- (Lin et al., 2007)
3- (Konishi et al., 2006)
4- (Sweeney et al., 2006)
5- (Wang et al., 1995)
6- (Yamanaka et al., 2004)
7- (Bligh et al., 1998)
8- (Fan et al., 2006)
9- (Bradbury et al., 2005)
10- (Ashikari et al., 2005)
11- (Song et al., 2007)

In addition to the suite of traits that were altered during domestication, the evolutionary history of rice is not complete without considering the deep population structure within *O. sativa*. Ancient records from the Han dynasty in China recognized the existence of two distinguishable types of rice, which were referred to as Hsien and Keng (Chou, 1948; Ting, 1957). We now refer to these two genetically distinct varietal groups as *Indica* and *Japonica,* respectively. The two groups have several morphological and physiological differences despite being grown in overlapping geographical ranges today. While *Indica* and *Japonica* represent the deepest genetic differentiation within *O. sativa*, five major subpopulations are widely recognized (*indica*, *tropical japonica*, *temperate japonica*, *aus*, and *Group V* (formerly *aromatic*) (Glaszmann, 1987; Garris et al., 2005) (Box 3.2). This population structure adds further complexity to our understanding of the domestication process that led to cultivated rice.

BOX 3.2: Subpopulation structure in *O. sativa*

There are two genetically distinct varietal groups within *O. sativa*, referred to as *Indica* and *Japonica*. Traditionally, the two groups have been distinguished based on morphological characters, including grain shape, apiculus hair length, leaf color, or through biochemical assays for reaction to phenol and sensitivity to potassium chlorate (Kato et al., 1928; Oka, 1988). There are also numerous reproductive barriers between the modern *Indica* and *Japonica* groups (Harushima et al., 2002). Yet the range of variation for any one of these phenotypic traits overlaps between the two groups, leading to confusion regarding the classification of particular genotypes (Oka, 1988).

Since publication of the first rice genetic map (McCouch et al., 1988), DNA markers have been widely used to explore the genetic architecture of rice. RFLP markers readily detect the differentiation between *Indica* and *Japonica* (Wang and Tanksley, 1989; Zhang et al., 1992). Isozymes, simple sequence repeat (SSR) markers, and single nucleotide polymorphisms (SNPs) provide a higher resolution of population structure. Using 169 SSR markers on a set of 234 diverse *O. sativa* genotypes, Garris et al. identified 5 subpopulations: *indica*, *aus*, *tropical japonica*, *temperate japonica*, and *aromatic* (Garris et al., 2005) (Figure 3.2). The same groups were identified using SNP markers derived from 111 randomly sequenced regions of the genome on a subset of 72 accessions (Caicedo et al., 2007). These studies were both consistent with the original study by Glaszmann (Glaszmann, 1987) using 15 isozyme markers on a larger set of 1700 diverse *O. sativa* genotypes. Of these five subpopulations, *indica* and *aus* belong to the *Indica* varietal group while *tropical japonica*, *temperate japonica*, and *aromatic* (now referred to by its isozyme group name, *Group V*) are closely related to the *Japonica* varietal group. A high degree of differentiation exists between the groups, with pairwise estimates of F_{st} ranging from 0.20 (*tropical* vs. *temperate japonica*) to 0.45 (*indica* vs. *temperate japonica*) (Garris et al., 2005).

The results of these studies highlight the importance of utilizing molecular markers to assign genotypes to the appropriate subpopulation for genetic analyses. For example, the Basmati rices (known for their aroma and specialized grain quality) have traditionally been classified as members the *Indica* varietal group due to their long, slender grain shape. However, the Basmati rice genotypes form a genetically distinct subpopulation (*Group V*) that is more closely related to *Japonica* than *Indica* (F_{st} for *tropical japonica-Group V* = 0.23; F_{st} for *indica-Group V* = 0.39) (Garris et al., 2005). Similarly, the *aus* varieties have traditionally been referred to as *indica* for lack of distinguishing morphological differences, but they are genetically divergent from *indica* (F_{st} for *indica-aus* = 0.26) (Garris et al., 2005).

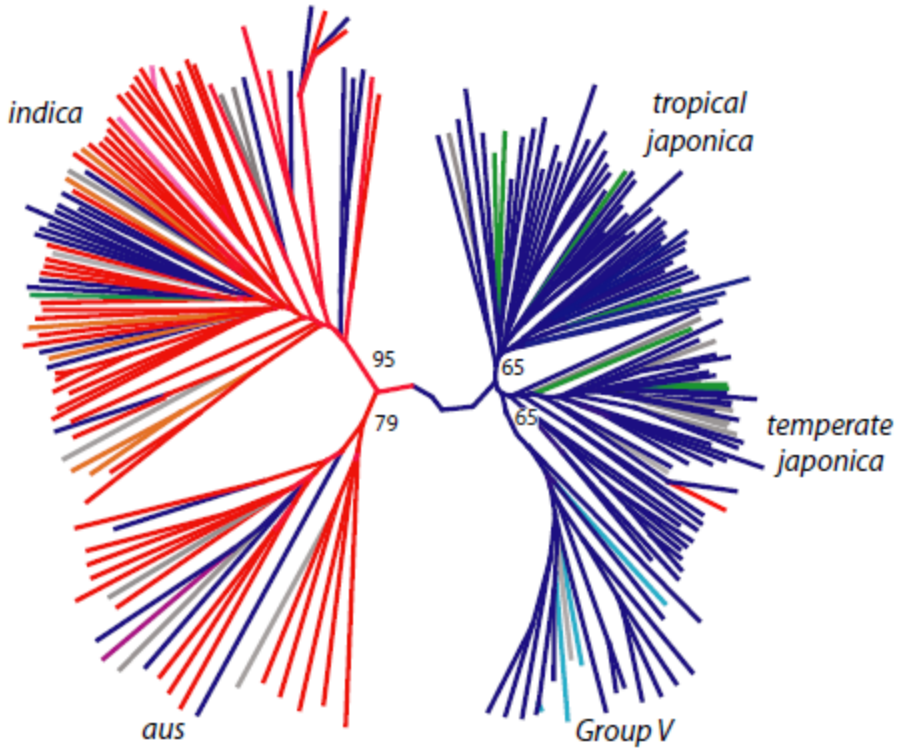


Figure 3.2: Subpopulation structure in *O. sativa*. *O. sativa* is characterized by the presence of deep genetic differentiation. This unrooted phylogenetic tree was constructed from data using 169 nuclear SSR and two chloroplast markers on 234 landraces of *O. sativa* (Garris et al., 2005). The branch structure reflects the phylogenetic relationships based on the nuclear SSR markers. The branch color corresponds to the chloroplast haplotype of each accession. This tree clearly illustrates the major division between the two varietal groups (*Indica* and *Japonica*), which are further subdivided into the five rice subpopulations: *indica*, *aus*, *tropical japonica*, *temperate japonica*, and *Group V*. (Reproduced and modified with permission from (Garris et al., 2005) ©2005 Genetics Society of America).

Modern genetics provides us with the necessary tools to understand the dynamics of the domestication process and to identify the key genes and alleles that were the targets of selection during rice domestication. Although only a limited number of genetic loci affecting key domestication traits in rice have been characterized at the gene level, an examination of the evolutionary history of these domestication genes provides novel insights into the movement of both cultivated rice and human populations across Asia. This review will critically examine the existing hypotheses used to describe the rice domestication process by integrating previous studies of genome-wide variation in rice with recent data on the evolutionary history of key domestication genes. The view that emerges is one that challenges us to rethink old paradigms as we continue to unravel the complex story of rice domestication.

Prevailing hypotheses regarding rice domestication

Several major hypotheses describe the process of rice domestication, each attempting to account for the evolution of the deeply differentiated subpopulations of *O. sativa*. First, a number of researchers have suggested that the *Indica* varietal group was originally domesticated from *O. rufipogon* and the *Japonica* group was derived later from *Indica* (Ting, 1957; Chang, 1976; Lu et al., 2002a). The greater genetic diversity found within *Indica* compared to *Japonica* (Garris et al., 2005) and the close genetic relationship between *Indica* and existing populations of both annual and perennial forms of *O. rufipogon* (Morishima and Gadrinab, 1987) provided the evidence to support this hypothesis. The second theory is that the *Indica-Japonica* differentiation occurred as a result of adaptation to different ecological and geographical environments following a single domestication of *O. sativa* from *O. rufipogon* (Oka and Morishima, 1982; Wang et al., 1984). A third hypothesis, originally proposed by Chou (Chou, 1948), is that rice was domesticated independently at least twice from a

pre-differentiated ancestral *O. rufipogon* gene pool (Second, 1982; Dally and Second, 1990; Nakano et al., 1992; Sun et al., 2002; Garris et al., 2005; Londo et al., 2006). An understanding of both the patterns of genome-wide variation in wild and cultivated rice and the number and origins of alleles associated with rice domestication genes will enable us to consider the validity of the proposed domestication hypotheses. According to the first hypothesis, if *Japonica* were derived from *Indica*, we would expect *Japonica* genotypes to be more closely related to *Indica* than to the *O. rufipogon* ancestor. Additionally, under this hypothesis we would expect to see a common set of domestication alleles in both groups with evidence that the alleles had originated in *Indica* and were later introgressed into *Japonica*. In the second scenario, if a single domestication from *O. rufipogon* were followed by a post-domestication divergence between *Indica* and *Japonica,* we would expect sequence differences between *Indica* and *Japonica* to post-date the time of domestication. Again under this theory, a common set of domestication alleles would be shared between the two groups. According to the third hypothesis, if *Indica* and *Japonica* were domesticated independently from a pre-differentiated *O. rufipogon* ancestor, we would expect to see sequence differences between the two groups to predate the time of domestication. We would also expect to see unique sets of *Indica*-specific and *Japonica*-specific domestication alleles originating from genetically distinct *O. rufipogon* populations. These group-specific alleles would each contribute to the same vital domestication phenotypes in their respective varietal groups. It is highly unlikely that an identical mutation would have arisen and been selected for in these two gene pools had they been independently domesticated.

Evolutionary history of genome-wide variation in *O. rufipogon* and *O. sativa*

To address these hypotheses, various studies have examined *O. rufipogon* for evidence of differentiation into *Indica* and *Japonica* types. Genetic analysis of *O. rufipogon* chloroplasts revealed two distinct haplotypes, which have clear frequency differences between the *Indica* and *Japonica* groups of *O. sativa* (Chen et al., 1993; Londo et al., 2006). Support for a differentiated *O. rufipogon* gene pool also comes from multiple studies using isozymes, restriction fragment length polymorphisms (RFLPs), single nucleotide polymorphisms (SNPs), transposable elements and the published genomic sequence, all demonstrating that *Indica* and *Japonica* accessions are more closely related to certain *O. rufipogon* accessions than to each other (Second, 1982; Lu et al., 2002a; Cheng et al., 2003; Hu et al., 2006; Londo et al., 2006; Caicedo et al., 2007; Duan et al., 2007; Rakshit et al., 2007).

The diverse geography and ecology of the broad region that is home to *O. rufipogon* (East, South and Southeast Asia) suggests that this ancestral species was broadly adapted to diverse climates and would have provided ample genetic diversity from which to have domesticated *O. sativa* on numerous occasions. Studies exploring the diversity within *O. rufipogon* have recently reported genetically identifiable population sub-structure with at least some hint of geographic associations (Sun et al., 2002; Zhou et al., 2003; Londo et al., 2006; Caicedo et al., 2007). The degree of genetic differentiation between populations of *O. rufipogon* increases with geographical distance, suggesting that geographical isolation played a major role in establishing the pre-differentiated gene pools within *O. rufipogon* (Zhou et al., 2003). Also, the inbreeding habit that accompanied rice domestication and transformed the highly outcrossing *O. rufipogon* (outcrossing rate of 30-50%) (Oka, 1988; Barbier, 1989) to the almost exclusively inbreeding *O. sativa* (outcrossing rate ~2%) (Oka,

1988; Messeguer et al., 2004) would have further promoted isolation between rice subpopulations.

A further source of evidence supporting multiple domestications comes from recent work estimating the divergence time between *Indica* and *Japonica*. The availability of complete genomic sequence for both *Japonica* and *Indica* (Goff et al., 2002; Yu et al., 2002; IRGSP, 2005) has allowed the estimation of the divergence time between these two groups. While the level of divergence varies across the genome (Zhu and Ge, 2005), two independent groups used a molecular clock approach to calculate the *Indica-Japonica* divergence time to be around 400,000 years ago (0.4 million years ago (mya)) (Ma and Bennetzen, 2004; Vitte et al., 2004). Another study evaluating the patterns of retrotransposon insertion in the two varietal groups estimated the divergence time to be at least 200,000 years ago (0.2 mya) (Tang et al., 2006). These dates are several orders of magnitude earlier than the first archaeological evidence for rice consumption by humans (Box 3.3), providing strong evidence that the divergence of the two cultivated rice gene pools predated rice domestication. Taken together, these studies contradict the first two hypotheses and support multiple domestications of *O. sativa* from a pre-differentiated *O. rufipogon* ancestor.

BOX 3.3: Effects of geography and climate on rice diversity

As early as the late 19th Century, De Candolle suggested that rice cultivation began in China, where the oldest historical records could be found (DeCandolle, 1882). By contrast, Vavilov designated India as the location of rice domestication (Vavilov, 1926). Since then, countless studies have generally agreed that the birthplace of rice domestication lies within a broad arc extending from the foothills of the Himalayas in eastern India through northern Burma, Thailand and Vietnam to southwest China (Chang, 1976; Oka, 1988). More recently, archaeological studies have strongly supported the middle Yangzte River Valley in China as the site where rice was first utilized by humans (Figure 3.3). The rice remains discovered at these sites indicate that rice was being gathered by humans, but it is unclear whether these remains represent foraged wild forms or truly domesticated forms of rice (Fuller, 2007). Analysis of phytolith samples (silicon fossils of rice plant cells) from the Diaotonghuan site in northern Jiangxi province (China) revealed the earliest known samples of domesticated rice, based on a morphological indicator (double-peaked glume cell phytoliths), that dated to 9,000-10,000 years ago (Zong et al., 2007).

Interestingly, despite this extensive archaeological evidence supporting the Yangtze River Valley in China as a region where rice was domesticated, the wild *O. rufipogon* populations in this region today exhibit very low genetic diversity. An examination of twelve wild rice populations in China demonstrated that the lowest genetic diversity existed in the northern-most populations near the Yangtze River (Zhou et al., 2003). It is also well known that the *Japonica* varietal group, believed to have undergone domestication in this region, has dramatically less genetic diversity than the *Indica* group (Garris et al., 2005). The reasons for the lower genetic diversity of *Japonica* relative to *Indica* may be explained by reviewing the history of the region. A major short-term climate change during a period known as the Younger Dryas, resulted in a return to glacial-like conditions across northern Asia from 13,000 to 11,500 years ago (Zhao, 1998; Lu et al., 2002b). The harsher climatic conditions likely precipitated a more rapid movement toward domestication of the *Japonica* gene pool as people were forced to cultivate rather than to rely on natural populations (Fuller, 2007). The colder climate would have eliminated a large portion of the *Japonica*-like wild ancestors in this region, limiting the gene flow between wild *Japonica*-like *O. rufipogon* and early *Japonica* domesticates. In contrast, it is likely that in the warmer tropics of South and Southeast Asia, the *Indica* gene pool was cultivated or foraged by humans for a longer time prior to domestication. During this time, it would have hybridized freely with diverse wild relatives, allowing *Indica* to accumulate higher levels of diversity than its *Japonica* counterpart.

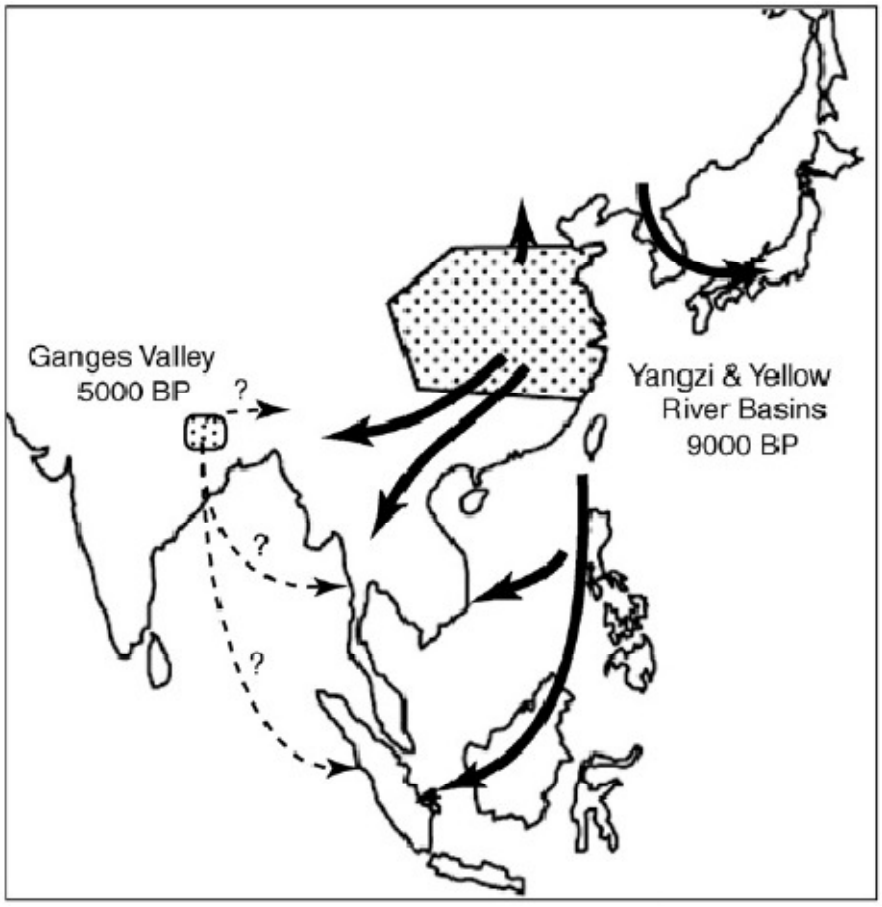


Figure 3.3: The origin and dispersal of cultivated rice. The majority of archaeological evidence points to the Yangtze River Valley of China as the birthplace of rice cultivation by humans. This region is where the oldest archaeological remains of rice have been uncovered. Radiocarbon data from fossilized rice remains from over 100 sites along the length of the Yangtze River were analyzed with the oldest samples coming from the middle Yangtze in Hubei and Hunan provinces, dating to around 11,500 years ago (Normile, 1997). Other studies analyzing radiocarbon data from fossilized rice remains in this region report dates ranging from 8,000 to 13,900 years ago (Jiang and Liu, 2006; Zong et al., 2007). The early rice domesticates are then thought to have moved north into Korea and Japan, and south and west into Southeast Asia (Diamond and Bellwood, 2003; Bellwood, 2005). Fuller (Fuller, 2002) postulates the Ganges Valley of India as a site of independent rice cultivation and potentially domestication. The dates for the earliest rice cultivation in this region based on archaeological evidence are more recent (around 5000 years ago). This region may represent the site of early cultivated forms of *Indica*-like rice. (Reproduced and modified with permission from P. Bellwood).

Evolutionary history of cloned rice domestication genes

To determine if the history of rice domestication alleles is consistent with multiple domestications, we will now examine several examples of cloned rice domestication genes whose evolutionary histories have been recently explored. If *Indica* and *Japonica* were independently domesticated as the patterns of genome-wide variation suggest, we would expect to see different domestication alleles present in the different varietal groups.

Grain Shattering: *Sh4*

One of the most universal changes during the domestication of the cereal crops was a reduction in grain shattering. This change was essential for humans to efficiently harvest the crop, but rendered the plant almost entirely dependent on human dispersal (Fuller, 2007). In a cross between an *indica* cultivar and an annual form of *O. rufipogon*, Li et al. identified a major QTL, *shattering4* (*sh4*), which affects grain shattering (Li et al., 2006a). The *sh4* gene was subsequently cloned and although its exact function remains unknown, *Sh4* it is thought to be a transcription factor involved in cell wall degradation and/or establishment of the abscission layer that releases the grain from the panicle (Li et al., 2006b; Lin et al., 2007). The functional nucleotide polymorphism in the gene was identified as a SNP causing a single amino acid change within the predicted DNA binding domain of SH4.

The *sh4* mutation is found in all five *O. sativa* subpopulations, whereas the dominant *Sh4* allele exists in all wild, shattering accessions of *O. rufipogon* along with six other closely related species of *Oryza* (Li et al., 2006a). Lin et al. recently examined *sh4* (which they refer to as *sha1*) in 96 *indica* and 112 *japonica* cultivars and found that they all contained the same functional SNP associated with the domesticated allele

(Lin et al., 2007). Additionally, *sh4* cDNA sequencing revealed an identical allele in four *Indica* and four *Japonica* accessions. The authors suggest that these results indicate the non-shattering *sh4* allele arose and was selected prior to the differentiation of *indica* and *japonica* from a common *O. rufipogon* ancestor. Although this is a logical conclusion, an alternative hypothesis is that the domesticated *sh4* allele arose and was selected during the domestication of one of the *O. sativa* subpopulations and was subsequently introgressed into all rice genotypes. To clarify the origin and routes of global dispersal of the non-shattering *sh4* allele, it will be necessary to examine the haplotype structure around the *sh4* gene in diverse populations of *O. rufipogon* and *O. sativa*.

Grain Shattering: *qSH1*

In addition to the *Sh4* gene, a second major shattering QTL, *qSH1*, was recently identified from an intra-specific cross between *temperate japonica* (cultivar Nipponbare) and *aus* (cultivar Kasalath) (Konishi et al., 2006). The functional mutation was a regulatory SNP 12 kilobases upstream of a BEL1-type homeobox transcription factor, which decreased expression of the transcription factor only at the provisional abscission layer, resulting in reduced shattering.

By examining the haplotypes around the *qSH1* gene in 118 rice lines including five *O. rufipogon* accessions, Konishi et al. deduced the subpopulation origin of this allele (Konishi et al., 2006). Two clear haplotype groups were found, corresponding to the *Indica* and *Japonica* varietal groups. The *Japonica* haplotypes carrying the non-shattering allele were most closely related to shattering *O. rufipogon* accession W1943, which originated in China and grouped closely with *Japonica* cultivars from a previous study (Cheng et al., 2003). This led to the conclusion that the causative

mutation in the domesticated *qSH1* allele occurred in early domesticates of the *Japonica* varietal group (Konishi et al., 2006). The non-shattering allele was prevalent in the *temperate japonica* subpopulation, but was absent in accessions from *tropical japonica* and *indica*, suggesting that this allele was not widely disseminated following its fixation within a portion of the *temperate japonica* gene pool. This distribution pattern is also consistent with the finding that *temperate japonica* varieties tend to be more difficult to thresh than varieties from other subpopulations, reflecting the presence of a non-shattering allele in addition to *sh4*.

Pericarp Color: *Rc*

All wild *Oryza* species and some of the early *O. sativa* landraces exhibit red pigmentation of the seed coat. However, virtually all modern rice cultivars lack red pigmentation and appear white or beige. Classical genetic analysis identified a single locus in rice, *Rc,* that conditioned the change from red to white pericarp (Kato and Ishikawa, 1921). The *Rc* gene was recently cloned and shown to encode a basic helix-loop-helix (bHLH) transcription factor (Sweeney et al., 2006). The *rc* allele differs from the wild type allele by a 14 base pair (bp) deletion that results in truncation of the protein before the bHLH domain and causes the shift from red to white pericarp (Sweeney et al., 2006; Furukawa et al., 2007).

To trace the origin and distribution of the *rc* allele, Sweeney et al. (Sweeney et al., 2007) evaluated 440 geographically and genetically diverse rice cultivars, representing landraces and modern varieties from all five of the subpopulations defined by Garris et al. (2005) (Box 3.2). They determined that 98% (330/337) of rice accessions with white pericarp contained the 14-bp deletion, whereas the mutation was not found in any of the landrace varieties with red pericarp (n = 103). To determine the

subpopulation origin of the 14-bp mutation, Sweeney et al. (Sweeney et al., 2007) used diverse red-grained varieties to identify polymorphisms in a 6.5 kilobase (kb) region around the *Rc* gene whose alleles were specific to either red *Indica* or red *Japonica*. They observed clear haplotype differentiation between *Indica* and *Japonica* across the region, making it possible to trace the origin of the *rc* allele conferring white pericarp. The white *rc* haplotype differed from the ancestral *Japonica* haplotype by only the 14-bp deletion (Figure 3.4). This analysis suggests that the mutation associated with the change from red to white pericarp in 98% of the *O. sativa* varieties surveyed originated in a *Japonica* ancestor (Sweeney et al., 2007). This study presents clear genetic evidence for a domestication allele arising in the *Japonica* varietal group and subsequently spreading across genetic, geographic and ethnic boundaries to become fixed in all cultivated rice.

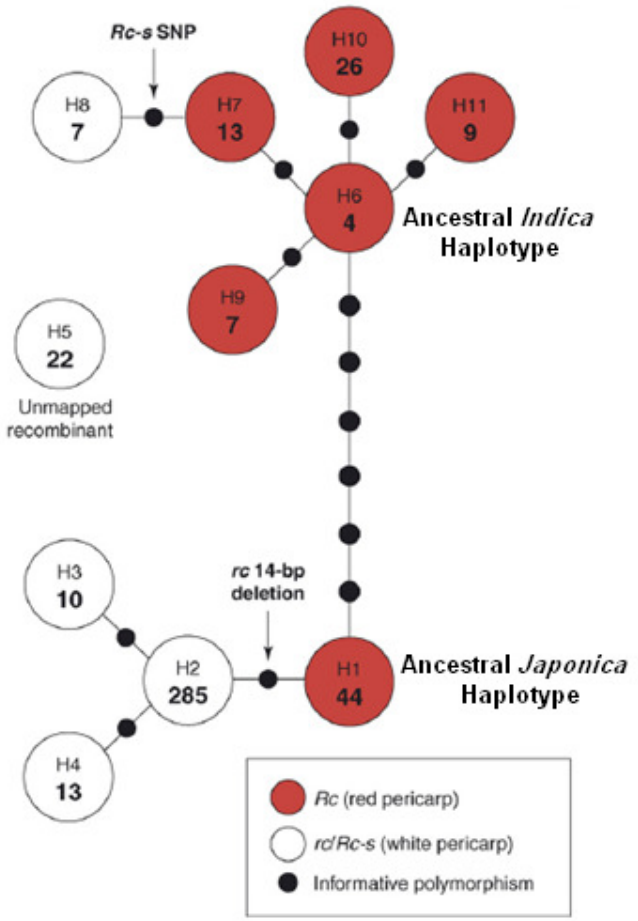


Figure 3.4: Haplotype network for the *Rc* gene (based on data from (Sweeney et al., 2007)). The *Rc* gene encodes a bHLH transcription factor that regulates the accumulation of red pigments in the rice pericarp. While genotypes with a functional *Rc* have red pericarp, genotypes with either a 14-bp deletion (creating the *rc* allele) or a SNP (creating the *Rc-s* allele) have white pericarp (Sweeney et al., 2006). In order to determine the origins and dispersal of these mutations in *Rc*, the sequences of 440 diverse rice cultivars were evaluated at this locus (Sweeney et al., 2007). This Figure uses the informative polymorphisms found around *Rc* to create a minimum spanning haplotype network. The open circles represent haplotypes H1-H11 and the numbers indicate the number of accessions in that haplotype. Circles with red shading indicate haplotypes with a functional *Rc* allele (red pericarp) while white shading signifies a nonfunctional *rc* or *Rc-s* allele (white pericarp). The two mutations in *Rc* resulting in white pericarp are denoted by arrows. Closed black circles represent the number of informative polymorphisms between haplotypes. The ancestral *Japonica* (H1) and *Indica* (H6) haplotypes are indicated. Haplotype H5 is an intragenic recombinant between two mapped haplotypes, and therefore was not included in the network. This network illustrates that 98% of all white genotypes surveyed were derived from the ancestral *Japonica* haplotype (Sweeney et al., 2007).

Pericarp Color: *Rc-s*

While 98% of the rice accessions with white pericarp surveyed had the same 14-bp deletion, two *aromatic* and five *aus* varieties had white pericarp but did not contain the *rc* allele. These seven varieties were sequenced to look for alternative mutations in the *Rc* gene that could explain the lack of pigment (Sweeney et al., 2007). A comparison of the sequences from these seven varieties to nine control varieties with red pericarp (from all five subpopulations) identified a single SNP in close proximity to the site where the 14-bp deletion was located in all the other white accessions. This second mutation introduces a premature stop codon that truncates the protein prior to the bHLH domain, similar to the *rc* mutation (Sweeney et al., 2007). Interestingly, none of the plants contained a complete knock-out of this gene, suggesting that the protein is critical to plant function independent of its ability to promote red grain color. All seven individuals that contained the *Rc-s* allele had an identical haplotype, which differed from the ancestral (red pericarp) *aus* haplotype only at the functional SNP. Thus, it was concluded that the *Rc-s* allele, which was only found in *aus* and a small percentage of *Group V* accessions, had its origins in the *aus* subpopulation (Sweeney et al., 2007). This clearly demonstrates that different mutations in the same gene can arise in different subpopulations over the course of domestication. Both *rc* and *Rc-s* apparently became the targets of selection and are virtually indistinguishable at the phenotypic level. It is interesting to note that unlike the *rc* and *sh4* alleles, but similar to the non-shattering *qSH1* allele, the *Rc-s* allele was not widely disseminated during the domestication of *O. sativa*.

Glutinous Rice: *Waxy*

Glutinous or "sticky" rice is used as a staple food in Southeast Asia and has long-standing culinary and cultural importance in China, Japan, and Korea. *Temperate*

japonica rices are typically valued for their low amylose content, leading to the soft, sticky texture of grains that are easy to eat with chopsticks, while *indica* rices are preferred with firm grains that flake apart when cooked, a consequence of higher amylose content. The glutinous phenotype is due to a defect in the *Waxy* (*Wx*) gene, which encodes a granule-bound starch synthase responsible for amylose biosynthesis in the grain (Wang et al., 1995). There are two functional forms of *Wx* (non-glutinous), with a clear divergence between the two varietal groups in rice: Wx^a is predominant in *Indica* and Wx^b is predominant in *Japonica*. The Wx^a allele was fixed in all *O. rufipogon* accessions examined (Yamanaka et al., 2004), indicating that the Wx^a allele is the wild type and Wx^b is the derived form, having been selected during the domestication of *Japonica*.

Unlike the Wx^a allele, the Wx^b allele contains a SNP at the 5' splice site of the first intron, causing incomplete post-transcriptional processing and lower amylose production in these genotypes (Wang et al., 1995; Bligh et al., 1998). A second mutation in either Wx^a or Wx^b creates the *wx* allele, which is recessive and fully glutinous. Glutinous rice strains can have either Wx^a- or Wx^b-derived *wx* alleles, but the Wx^b-derived *wx* allele is predominant, as it was found in 97% of 353 glutinous samples from both varietal groups collected from diverse regions across Asia (Yamanaka et al., 2004). This finding demonstrates that the majority of glutinous rice cultivars from both *Indica* and *Japonica* carry the *Japonica*-derived allele, mirroring the situation described above for the *rc* mutation. Also, similar to the *Rc-s* allele, glutinous genotypes derived from the *indica* Wx^a allele were not widely disseminated.

Other domestication-related traits:

The examples presented above include only the rice domestication genes that have been cloned and investigated in terms of their evolutionary origins. Several other genes that affect domestication-related traits have been cloned in rice, including *GS3* and *GW2*, which condition grain size and weight (Fan et al., 2006; Song et al., 2007), *Gn1a*, associated with grain number (Ashikari et al., 2005), and *BADH2*, which conditions fragrance (Bradbury et al., 2005) (Table 3.1). Further examination of these genes will determine whether they were involved in the rice domestication process or have been the targets of more recent selection.

What we can interpret from the examples presented above is that several key domestication alleles had a single origin and subsequently became fixed across all rice subpopulations. In the case of white pericarp, the *rc* allele arose in the *Japonica* gene pool and was then introgressed into the other subpopulations (Sweeney et al., 2007). This story is repeated for the Wx^b–derived glutinous rice mutation (*wx*) that originated in the *Japonica* group and was also introgressed and dispersed (Yamanaka et al., 2004). In the future, we may find that the *sh4* allele for non-shattering followed a similar evolutionary path. By contrast, several other domestication alleles were not widely disseminated across subpopulations. The *Rc-s* allele for white pericarp is largely restricted to the *aus* subpopulation, with some introgression into the *Group V* subpopulation (Sweeney et al., 2007). The *qSH1* allele for non-shattering and the Wx^b non-glutinous allele both remain largely contained within the *temperate japonica* subpopulation (Yamanaka et al., 2004; Konishi et al., 2006).

Current observations regarding the history of rice domestication

The examples of rice domestication genes, coupled with the studies involving genome-wide variation in both *O. rufipogon* and *O. sativa*, cannot be reconciled with any of the three prevailing hypotheses about rice domestication. Available genetic data suggest that there were pre-differentiated gene pools within the *O. rufipogon* ancestor and these gene pools gave rise to the *Indica* and *Japonica* varietal groups. Yet, the presence of identical domestication alleles in both groups strongly suggests that these domestications were not entirely independent. Gene flow across the distinct gene pools of rice appears to have been crucial to the domestication process. Taken together, the current data support the recently proposed "combination model" for rice domestication (Sang and Ge, 2007). In this model, the early *Japonica* and *Indica* cultivars were domesticated from divergent *O. rufipogon* populations and subsequent introgression of key domestication alleles between these early domesticates resulted in a common set of domestication alleles being fixed in all modern varieties. While all of the evolutionary data from domestication genes thus far demonstrate gene flow from *Japonica* to *Indica*, only a limited number of rice domestication alleles have thus far been described. This situation is likely to change as the costs of high resolution genotyping and re-sequencing continue to decline, and as the development of new statistical approaches improves our ability to identify and characterize the targets of selection in diverse germplasm. Whether future genetic studies involving larger numbers of domestication genes will substantiate this observation of gene flow from *Japonica* to *Indica* or will identify ubiquitous domestication alleles arising in other subpopulations remains to be seen.

If we accept the "combination model" for rice domestication, how can we reconcile the maintenance of deep genetic differentiation between rice varietal groups with the

movement of domestication genes? A recent study involving SNP data on 20 accessions of *O. rufipogon* and 72 landrace and modern *O. sativa* cultivars demonstrated that *O. rufipogon* has the greatest level of heterozygosity, followed by the landraces, with very low levels observed in modern cultivars. These data are consistent with a gradual change from an outcrossing to an inbreeding habit during the domestication process, which coincided with the rise of the deeply differentiated subpopulations of *O. sativa* (Caicedo et al., 2007). Numerous reproductive barriers help maintain this genetic differentiation in modern *O. sativa* (Harushima et al., 2002), but cross-hybridization was likely to have been easier in early landrace varieties. Evidence of *Indica* x *Japonica* hybridization comes from the recovery of early domesticates containing *Indica*-like organellar genomes in *Japonica* nuclear backgrounds and vice versa throughout Himalayan region of Southeast Asia (Sun et al., 2002; Ishikawa et al., 2006). It therefore appears that higher outcrossing rates and fewer sterility barriers provided a natural corridor for domestication genes to be transferred between the two emerging varietal groups. Yet, despite evidence of substantial hybridization, the two *O. sativa* varietal groups remain largely isolated from each other at the genetic level. This apparent contradiction may be resolved by the recent discovery of well-defined regions of high and low divergence between *Indica* and *Japonica* in the rice nuclear genome (Zhu and Ge, 2005). Genomic regions of low diversity may be more permissive to recombination and therefore introgression between groups while regions of high genetic diversity may reinforce the sterility barriers and help maintain the independence of the *Indica* and *Japonica* gene pools (Zhu and Ge, 2005). Consistent with this hypothesis, Sweeney et al. recently documented that the domestication allele for white pericarp, *rc,* entered the *Indica* gene pool as a small (<1 Mb) introgression from a *Japonica* ancestor and that this region of introgression is characterized by an abrupt drop in F_{st} values (Sweeney et al.,

2007). This indicates that the region of introgression can be classified as an island of low divergence in a background of well differentiated *Indica* or *Japonica*-like genomes. Thus, observations at both the gene and the genome levels support the view that relatively small portions of genetic information were exchanged across gene pools during the process of rice domestication (Zhu and Ge, 2005; Sweeney et al., 2007). The introgressions containing key domestication loci would have been maintained by human selection. Limited amounts of introgression are consistent with both the deep genetic differentiation that distinguishes the *Indica* and *Japonica* varietal groups and the presence of common domestication alleles across *O. sativa*.

Concluding remarks – Lessons for the future

The story of rice domestication is one of both genetic exchange and genetic containment. It involved episodes of hybridization and introgression between ancient gene pools against a backdrop of emerging inbreeding and sterility barriers that progressively restricted gene flow between subpopulations. These competing evolutionary forces, sculpted by human selection, gave rise to an array of interconnected, but well-differentiated, subpopulations of *O. sativa.* Although these subpopulations are intimately associated with their specific geographies, ecologies and cultures, the history of rice domestication is a dynamic rather than a static story.

Today, these same forces of genetic exchange and containment are at work, and the tools of modern genetics provide new opportunities for identifying and mobilizing genetic variation. High yielding hybrid rice varieties take advantage of the traditional subpopulation structure to systematically exploit new combinations of genes by bringing together inbred parents from different subpopulations. The development of inter- and intra-specific introgression lines mimics the ancient road to domestication

by allowing only selected segments of the donor genome to be introgressed into adapted genetic backgrounds. Genomics techniques offer a rapid way of identifying useful variation, which can then be exploited with transgenic technology. Yet, with all of these possibilities for genetic manipulation available to the modern plant breeder, the most valuable source of genetic variation today is still the wild germplasm itself. Learning how humans once domesticated crops from wild ancestors may provide more than a retrospective. As we seek new ways of harnessing genetic variation in the context of the 21st century, understanding the dynamics of crop domestication may offer important lessons for the future.

REFERENCES

Ashikari M., Sakakibara H., Lin S., Yamamoto T., Takashi T., Nishimura A., Angeles E.R., Qian Q., Kitano H., Matsuoka M. (2005) Cytokinin oxidase regulates rice grain production. Science 309:741-745

Barbier P. (1989) Genetic variation and ecotypic differentiation in the wild rice species *O. rufipogon*. II. Influence of the mating system and life history traits on the genetic structure of populations. Japanese Journal of Genetics 63:273-285

Bellwood P. (2005) Examining the language/farming dispersal hypothesis in the East Asian context. In: Sagart L., Blench R., Sanchez-Mazas A. (eds) The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics. Routledge Curzon, London, pp 17-30

Bligh H.F.J., Larkin P.D., Roach P.S., Jones C.A., Fu H., Park W.D. (1998) Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose rice varieties. Plant Molecular Biology 38:407-415

Bradbury L.M.T., Fitzgerald T.L., Henry R.J., Jin Q., Waters D.L.E. (2005) The gene for fragrance in rice. Plant Biotechnology Journal 3:363-370

Caicedo A.L., Williamson S.H., Hernandez R.D., Boyko A., Fledel-Alon A., York T.L., Polato N.R., Olsen K.M., Nielsen R., McCouch S.R., Bustamante C.D., Purugganan M.D. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genetics 3:e163

Chang T.-T. (1976) The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica 25:425-441

Chang T.T. (1995) Rice. In: Smartt J.S., N.W. (ed) Evolution of Crop Plants. Longman Scientific and Technical

Chen W.B., Nakamura I., Sato Y.I., Nakai H. (1993) Distribution of deletion type in

cpDNA of cultivated and wild rice. Japanese Journal of Genetics 68:597-603

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67-75

Chou S.L. (1948) China is the place of origin of rice. J Rice Soc 7:53-54

Dally A.M., Second G. (1990) Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. Theoretical and Applied Genetics 80:209-222

DeCandolle A.P. (ed) (1882) Origins of Cultivated Species. Hafner

Diamond J., Bellwood P. (2003) Farmers and their languages: The first expansions. Science 300:597-603

Doebley J.F., Gaut B.S., Smith B.D. (2006) The molecular genetics of crop domestication. Cell 127:1309-1321

Duan S., Lu B., Li Z., Tong J., Kong J., Yao W., Li S., Zhu Y. (2007) Phylogenetic analysis of AA-genome *Oryza* species (*Poaceae*) based on chloroplast, mitochondrial, and nuclear DNA sequences. Biochemical Genetics 45:113-129

Fan C., Xing Y., Mao H., Lu T., Han B., Xu C., Li X., Zhang Q. (2006) *GS3* , a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics 112:1164-1171

Fuller D. (2002) An agricultural perspective on Dravidian historical linguisitics: archaeological crop packages, livestock and Dravidian crop vocabulary. In: Bellwood P., Renfrew C. (eds) Examining the Farming/Lanuage Dispersal Hypothesis. McDonald Institute for Archaeological Research, pp 191-213

Fuller D.Q. (2007) Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World. Annals of Botany

100:903-924

Furukawa T., Maekawa M., Oki T., Suda I., Iida S., Shimada H., Takamure I., Kadowaki K.-i. (2007) The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. The Plant Journal 49:91-102

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638

Gepts P. (2004) Crop domestication as a long-term selection experiment. Plant Breeding Reviews 24:1-44

Glaszmann J.C. (1987) Isozymes and classification of Asian rice varieties. Theoretical and Applied Genetics 74:21-30

Goff S.A., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100

Gu X.-Y., Kianian S.F., Foley M.E. (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). Genetics 166:1503-1516

Gu X.-Y., Kianian S.F., Foley M.E. (2005a) Phenotypic selection for dormancy introduced a set of adaptive haplotypes from weedy into cultivated rice. Genetics 171:695-704

Gu X.-Y., Kianian S.F., Hareland G.A., Hoffer B.L., Foley M.E. (2005b) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theoretical and Applied Genetics 110:1108-1118

Harushima Y., Nakagahra M., Yano M., Sasaki T., Kurata N. (2002) Diverse variation of reproductive barriers in three intraspecific rice crosses. Genetics 160:313-322

Hu H., Mu J., Zhang H.-J., Tao Y.-Z., Han B. (2006) Differentiation of a miniature inverted transposable element (MITE) system in Asian rice cultivars and its

inference for a diphyletic origin of two subspecies of Asian cultivated rice. Journal of Integrative Plant Biology 48:260-267

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793-800

Ishikawa R., Yamanaka S., Fukuta Y., Chitrakon S., Bounphanousay C., Kanyavong K., Tang L., Nakamura I., Sato T., Sato Y.I. (2006) Genetic erosion from modern varieties into traditional upland rice cultivars. Genetic Resources and Crop Evolution 53:245-252

Jiang L., Liu L. (2006) New evidence for the origins of sedentism and rice domestication in the Lower Yangzi River, China. Antiquity 80:355-361

Kato S., Ishikawa J. (1921) On the inheritance of the pigment of red rice. Japanese Journal of Genetics 1:1-7

Kato S., Kosaka H., Hara K. (1928) On the affinity of rice varieties as shown by fertility of hybrid plants. Bull Sci Fac Agric Kyushu Univ 3:132-147

Khush G.S. (1997) Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25-34

Kinoshita T. (1998) Linkage mapping using mutant genes in rice. Rice Genetics Newsletter 15:13-74

Konishi S., Izawa T., Lin S.Y., Ebana K., Fukuta Y., Sasaki T., Yano M. (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392-1396

Li C., Zhou A., Sang T. (2006a) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytologist 170:185-194

Li C., Zhou A., Sang T. (2006b) Rice domestication by reducing shattering. Science 311:1936-1939

Lin Z., Griffith M., Li X., Zhu Z., Tan L., Fu Y., Zhang W., Wang X., Xie D., Sun C. (2007) Origin of seed shattering in rice (*Oryza sativa* L.). Planta 226:11-20

Londo J.P., Chiang Y.-C., Hung K.-H., Chiang T.-Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proceedings of the National Academy of Sciences 103:9578-9583

Lu, Lu B.R., Zheng, Zheng K., Qian, Qian H., Zhuang, Zhuang J. (2002a) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics 106:101-106

Lu H., Liu Z., Wu N., Berne S., Saito Y., Liu B., Wang L. (2002b) Rice domestication and climatic change: phytolith evidence from East China. Boreas 31:378-385

Ma J., Bennetzen J.L. (2004) Rapid recent growth and divergence of rice nuclear genomes. Proceedings of the National Academy of Sciences of the United States of America 101:12404-12410

McCouch S.R., Kochert G., Yu Z.H., Wang Z.Y., Khush G.S., Coffman W.R., Tanksley S.D. (1988) Molecular mapping of rice chromosomes. Theoretical and Applied Genetics 76:815-829

Messeguer J., Marfà V., Català M.M., Guiderdoni E., Melé E. (2004) A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. Molecular Breeding 13:103-112

Miyata M., Yamamoto T., Komori T., Nitta N. (2007) Marker-assisted selection and evaluation of the QTL for stigma exsertion under *japonica* rice genetic background. Theoretical and Applied Genetics 114:539-548

Morishima H. (2001) Evolution and domestication of rice. Rice Genetics Newsletter IV:63-78

Morishima H., Gadrinab L.U. (1987) Are the Asian common wild rices differentiated into the *indica* and *japonica* types? In: Hsieh S.C. (ed) Crop Exploration and Utilization of Genetic Resources. Taichung District Agricultural Station, pp

11-20

Nakano M., Yoshimura A., Iwata N. (1992) Phylogenetic study of cultivated rice and its wild relatives by RFLP. Rice Genetics Newsletter 9:132-134

Normile D. (1997) Archaeology: Yangtze seen as earliest rice site. Science 275:309-300

Oka H.-I., Morishima H. (1982) Phylogenetic differentiation of cultivated rice, XXIII. Potentiality of wild progenitors to evolve the *Indica* and *Japonica* types of rice cultivars. Euphytica 31:41-50

Oka H.I. (1988) Origin of Cultivated Rice. Elsevier Science/Japan Scientific Societies Press, Tokyo

Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N., Terauchi R. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theoretical and Applied Genetics 114:731-743

Sang T., Ge S. (2007) The puzzle of rice domestication. Journal of Integrative Plant Biology 49:760-768

Second G. (1982) Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Japanese Journal of Genetics 57:25-57

Song X.-J., Huang W., Shi M., Zhu M.-Z., Lin H.-X. (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nature Genetics 39:623-630

Sun C., Wang X., Yoshimura A., Doi K. (2002) Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). Theoretical and Applied Genetics 104:1335-1345

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D., McCouch S.R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3:e133

Sweeney M.T., Thomson M.J., Pfeil B.E., McCouch S. (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294

Tang T., Lu J., Huang J., He J., McCouch S.R., Shen Y., Kai Z., Purugganan M.D., Shi S., Wu C.-I. (2006) Genomic variation in rice: Genesis of highly polymorphic linkage blocks during domestication. PLoS Genetics 2:e199

Thomson M.J., Tai T.H., McClung A.M., Lai X.H., Hinga M.E., Lobos K.B., Xu Y., Martinez C.P., McCouch S.R. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. Theoretical and Applied Genetics 107:479-493

Ting Y. (1957) The origin and evolution of cultivated rice in China. Acta Agron Sinica 8:243-260

Uga Y., Fukuta Y., Cai H.W., Iwata H., Ohsawa R., Morishima H., Fujimura T. (2003) Mapping QTLs influencing rice floral morphology using recombinant inbred lines derived from a cross between *Oryza sativa* L. and *Oryza rufipogon* Griff. Theoretical and Applied Genetics 107:218-226

Vavilov N.I. (1926) Studies in the origin of cultivated plants. Bull Appl Bot 16:139-248

Vitte C., Ishii T., Lamy F., Brar D., Panaud O. (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Molecular Genetics and Genomics 272:504-511

Wang X.K., Cheng K.S., Lu Y.X., Luo J., Huang L.W., Liu G.R. (1984) Coordinated

studies on genetic resources of rice in Yunnan. B. Glabrous hull rice (guangkedao) in Yunnan. Res Bull Beijing Agric Univ 10:333-344

Wang Z.-Y., Zheng F.-Q., Shen G.-Z., Gao J.-P., Snustad D.P., Li M.-G., Zhang J.-L., Hong M.-M. (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. The Plant Journal 7:613-622

Wang Z., Tanksley S.D. (1989) Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32

Xiong L.Z., Liu K.D., Dai X.K., Xu C.G., Zhang Q. (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theoretical and Applied Genetics 98:243-251

Yamanaka S., Nakamura I., Watanabe K., Sato Y.-I. (2004) Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theoretical and Applied Genetics 108:1200-1204

Yu J., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92

Zhang Q., Maroof M.A.S., Lu T.Y., Shen B.Z. (1992) Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theoretical and Applied Genetics 83:495-499

Zhao Z. (1998) The Middle Yangtze region in China is one place where rice was domesticated: phytolith evidence. Antiquity 72:885

Zhou H.-f., Xie Z.-w., Ge S. (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. Theoretical and Applied Genetics 107:332-339

Zhu Q., Ge S. (2005) Phylogenetic relationships among A-genome species of the

genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytologist 167:249-265

Zhu Q., Zheng X., Luo J., Gaut B.S., Ge S. (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. Molecular Biology and Evolution 24:875-888

Zong Y., Chen Z., Innes J.B., Chen C., Wang Z., Wang H. (2007) Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. Nature 449:459-462

CHAPTER 4:
THE ORIGIN AND EVOLUTION OF FRAGRANCE IN RICE (*ORYZA SATIVA* L.)

ABSTRACT

Fragrance in the grain is one of the most highly-valued grain quality traits in rice, yet the origin and evolution of the betaine aldehyde dehydrogenase gene (*BADH2*) underlying this trait remains unclear. In this study, we identify eight novel and putatively non-functional alleles of the *BADH2* gene, and show that these alleles have distinct geographic and genetic origins. Despite multiple origins of the fragrance trait, a single allele, *badh2.1*, is the predominant allele in virtually all fragrant rice varieties today, including the widely recognized Basmati and Jasmine types. Haplotype analysis allowed us to establish a single origin of the *badh2.1* allele within the *Japonica* varietal group and demonstrate the introgression of this allele from *Japonica* to *Indica*. Basmati-like accessions were nearly identical to the ancestral *Japonica* haplotype across a 5.3 megabase region flanking *BADH2* regardless of their fragrance phenotype, demonstrating a close evolutionary relationship between Basmati varieties and the *Japonica* gene pool. These results clarify the relationships among fragrant rice varieties and challenge the traditional assumption that the fragrance trait arose in the *Indica* varietal group.

INTRODUCTION

Recent findings suggest that although Asian cultivated rice (*Oryza sativa*) is comprised of several genetically distinct groups, a number of the alleles responsible for key domestication and grain quality traits are shared among these groups. Given the complex evolutionary history of rice, the origin of these genetic alterations and

how they came to exist across the highly divergent subpopulations of *O. sativa* remain central questions in rice evolutionary biology.

Fragrance is considered one of the most important grain quality traits in rice, as it is a key factor in determining market price and is related to both local and national identity (Bhattacharjee et al., 2002; Fitzgerald et al., 2009). Investigations into the genetic basis of fragrance in rice led to the identification of a single locus on chromosome 8 (*fgr*) associated with fragrance (Kadam and Patankar, 1938; Jodon, 1944). Fine mapping (Ahn et al., 1992; Chen et al., 2006) and subsequent sequence analysis identified a betaine aldehyde dehydrogenase gene, *BADH2*, associated with the fragrant phenotype (nomenclature follows (McCouch & CGSNL, 2008)). The functional mutation creating the recessive *badh2.1* allele was first described as 3 single nucleotide polymorphisms (SNPs) and an 8 base pair (bp) deletion in the 7^{th} exon of the gene that resulted in a premature stop codon and putatively truncated the BADH2 protein (Bradbury et al., 2005b; McCouch and CGSNL, 2008). Other sequence alignments have been used to describe this complex mutation (Amarawathi et al., 2008; Bourgis et al., 2008), and so the mutation in *badh2.1* will hereafter be referred to as the functional nucleotide polymorphism (FNP). Recent surveys of diverse fragrant germplasm support the association of *badh2.1* with fragrance (Bourgis et al., 2008; Fitzgerald et al., 2008; Shi et al., 2008) and transformation of a fragrant variety with the dominant, non-fragrant allele has been shown to abolish fragrance (Chen et al., 2008), confirming that *BADH2* is the major genetic determinant of fragrance in rice.

Over 100 volatile compounds have been detected in fragrant rice varieties, but the major compound responsible for the characteristic aroma is 2-acetyl-1-pyrroline (2AP)

(Buttery et al., 1982; Paule and Powers, 1989). This compound, which is produced in all parts of the rice plant except the roots, has a very low odor threshold, allowing humans to detect it at minute concentrations in field-grown plants or crushed leaf tissue, as well as in the grain before, during and after cooking (Buttery et al., 1983). While the biochemical pathway leading to 2AP synthesis has not been fully resolved, it is believed the BADH2 protein catalyzes the oxidation of γ-aminobutyraldehyde (AB-ald; a 2AP precursor), so that a nonfunctional allele results in the accumulation of both AB-ald and its cyclic form, Δ^1pyrroline, resulting in enhanced 2AP synthesis (Bradbury et al., 2008; Chen et al., 2008).

Oryza sativa consists of two major varietal groups, *Indica* (*Hsien*) and *Japonica* (*Keng*) (capitalized when referring to varietal groups), that have been recognized in China since ancient times (Chou, 1948; Ting, 1957). A set of 15 isozyme markers was able to further subdivide the two major varietal groups into six genetically differentiated subpopulations corresponding to well-recognized ecotypes (Glaszmann, 1987). Subsequent studies using SSRs (Garris et al., 2005) and SNPs (Caicedo et al., 2007) distinguished five genetically defined groups that roughly corresponded to the isozyme groups identified by Glaszmann: *indica*, *aus*, *temperate japonica*, *tropical japonica*, and *aromatic* (lower case when referring to subpopulations) (Figure 4.1). Phylogenetic analysis and F_{ST} values demonstrate a close evolutionary relationship between the *aromatic*, *temperate japonica*, and *tropical japonica* subpopulations, which comprise the *Japonica* varietal group, while the *indica* and *aus* subpopulations have a distinct ancestry and are recognized as members of the *Indica* varietal group (Garris et al., 2005; Caicedo et al., 2007). Despite its name, the *aromatic* subpopulation is phenotypically diverse and includes both fragrant and non-fragrant

varieties. To avoid confusion, we will hereafter refer to the *aromatic* subpopulation by its isozyme name, *Group V* (Glaszmann, 1985).

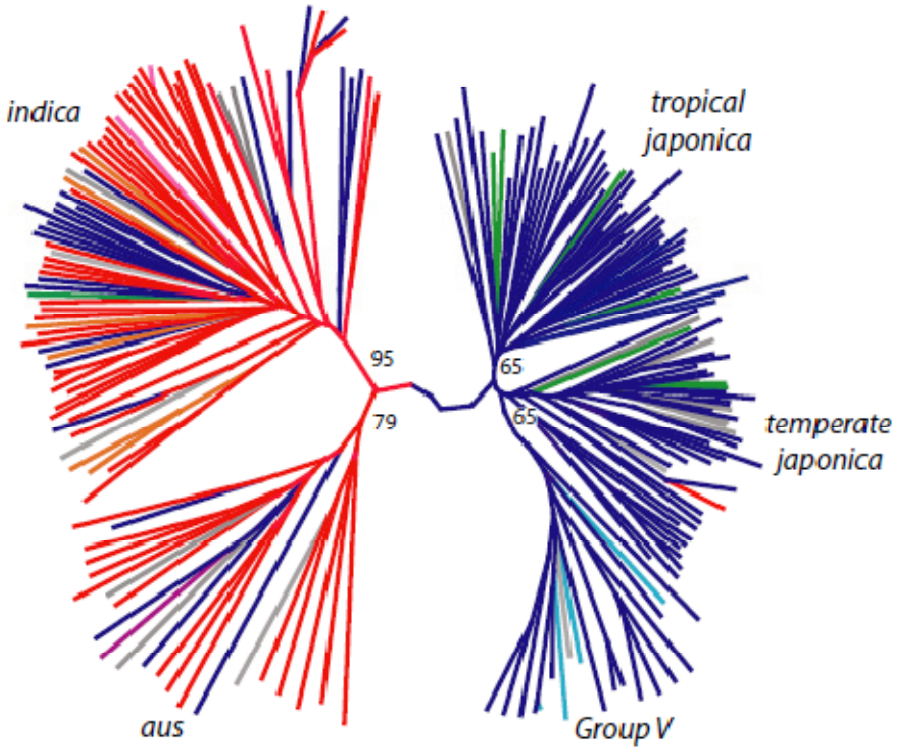


Figure 4.1: Subpopulation structure in *O. sativa.* Unrooted neighbor-joining tree constructed from 169 nuclear SSRs (Garris et al., 2005). Branch color corresponds to chloroplast haplotype. Bootstrap values (out of 100) are indicated at the branch points. This tree clearly illustrates the major division between the two varietal groups (*Indica* and *Japonica*), which are further subdivided into the five rice subpopulations: *indica*, *aus*, *tropical japonica*, *temperate japonica* and *Group V* (*aromatic)*. Reproduced and modified, with permission from (Garris et al., 2005).

Fragrant accessions have been identified within at least three of the distinct genetic subpopulations of rice, including *Group V* (i.e. "Basmati" and "Sadri" varieties), *indica* (i.e. "Jasmine" varieties), and *tropical japonica*. In surveys of diverse fragrant rice accessions, nearly all have been shown to carry the *badh2.1* allele, suggesting that this allele is common by descent in fragrant rice varieties (Bradbury et al., 2005b; Bourgis et al., 2008; Fitzgerald et al., 2008). Recently, a second mutation in the *BADH2* gene, *badh2.2*, was found to be associated with fragrance within a limited set of germplasm from China (Shi et al., 2008). Evidence that there may be additional mutations in the pathway leading to 2AP synthesis comes from a rigorous study involving a diverse panel of fragrant germplasm that identified several accessions, mostly from Southeast Asia, that had elevated levels of 2AP but did not carry the *badh2.1* allele (Fitzgerald et al., 2008). Given that a single allele is largely responsible for fragrance in rice, the goal of this study was to investigate the origin of this allele and trace its ancestry among the highly genetically divergent subpopulations of rice. We also set out to identify additional functional mutations in the *BADH2* gene that may be responsible for independent, local origins of the fragrant phenotype.

MATERIALS AND METHODS

Plant materials

Our germplasm panel consisted of a total of 280 *O. rufipogon* and 242 *O. sativa* accessions collected from 38 countries across Asia. We also obtained 26 fragrant accessions lacking the *badh2.1* allele from a previous study (Fitzgerald et al., 2008). A complete list of the accessions used in this study can be found in Supplemental Table 4.1. All *O. sativa* accessions not previously reported were analyzed with diagnostic SSR markers to determine their subpopulation identities.

DNA extraction, PCR, and sequencing

DNA extraction was performed using a potassium acetate-SDS protocol for leaf tissue (Dellaporta et al., 1983) and a modified protocol for milled seeds (Kang et al., 1998). The functional marker for *badh2.1* (Bradbury et al., 2005a) was used to genotype our germplasm panel. For gene haplotype analysis, eight ~700 bp amplicons were sequenced across the coding region of the *BADH2* gene, resulting in ~5 kb of aligned sequence. For extended haplotype analysis, 24 regions were sequenced, spanning 3.2 Mb upstream and 2.1 Mb downstream of *BADH2*, and resulting in over 13 kb of aligned sequence. A previously described MITE polymorphism (Bourgis et al., 2008) was also genotyped in the *O. sativa* panel. A complete list of the primers for the 32 sequenced regions can be found in Supplemental Table 4.5. PCR products were purified and sequencing was performed on ABI Prism 3700/3100 DNA analyzers (Applied Biosystems, Foster City, CA) at the Cornell Life Sciences Core Laboratories Center. Sequences were aligned using the CodonCode Aligner program (CodonCode, Dedham, MA) and the ends of amplicons were trimmed to remove low quality sequences. Singletons and ambiguous sites were re-sequenced as necessary. The nonsynonymous polymorphisms found in the fragrant accessions lacking any known nonfunctional allele at *BADH2* were re-sequenced several times for confirmation.

2AP phenotyping

Extraction of 2-acetyl-1-pyrroline with dichloromethane was performed using a modified method (52). Chemically synthesized 2AP was provided by Dr. T. Yoshihashi (Japan International Research Centre for Agricultural Sciences, Ibaraki, Japan) and was used to quantify 2AP in the samples. Each sample was extracted and analyzed on six different occasions, and from at least three biological replicates.

Haplotype and genetic diversity analysis across the *BADH2* region

Aligned sequences for the eight amplicons across the coding region of the *BADH2* gene in the 242 *O. sativa* accessions were imported into the TASSEL program to extract all polymorphisms present at a frequency above 5% in the sample for constructing gene haplotypes (Supplemental Table 4.2). The highly polymorphic SSR at position 20249280 was either long (TA_{11-14}) or short (TA_{6-8}), which corresponded to the *Japonica* and *Indica* varietal groups, respectively, and so was coded as "1" if the allele had greater than 8 repeats and "0" if the allele had 8 or fewer repeats. A total of eight gene haplotypes were inferred using 238 accessions, with four recombinant haplotypes not included in Figure 4.2*A*. The Bayesian clustering program STRUCTURE was used to analyze the gene haplotypes and the highest likelihood was obtained at K=2 clusters, which were then labeled *Japonica* Gene Haplotype Cluster (Jap_GH) and *Indica* Gene Haplotype Cluster (Ind_GH).

For the extended haplotype analysis, we sequenced 24 regions flanking *BADH2* in the same 242 *O. sativa* accessions as described above (Supplemental Table 4.5). TASSEL was used to extract the 216 polymorphisms present at a frequency above 5% in the sample. To reduce the number of haplotypes, we applied the following criteria to select ancestrally informative polymorphisms (AIPs): a polymorphism must 1) have a significant frequency difference between the *Japonica* varietal group and the *indica* subpopulation ($p< 0.00003$) and 2) be present at a frequency above 20% in the population. A total of 78 polymorphisms from 16 of the sequenced amplicons, the FNP, and the MITE polymorphism met these criteria, and were used in Figure 4.2*B* and Supplemental Table 4.3.

Polymorphism data from the 24 sequenced regions flanking *BADH2* were used to calculate extended haplotype homozygosity (EHH) across the region in the 242 *O. sativa* accessions, following the method described previously (Sabeti et al., 2002).

RESULTS

Frequency of the *badh2.1* allele in diverse rice germplasm

We examined the occurrence of the *badh2.1* allele in 280 accessions of wild rice (*O. rufipogon*/*O. nivara*) and found that it was absent from all wild genotypes, except for a single accession that was heterozygous for the allele (Table 4.1). This wild accession exhibited several traits characteristic of domesticated rice, including white pericarp, suggesting it was the result of a recent hybridization event with a fragrant accession of *O. sativa*. A diverse collection of 176 *O. sativa* accessions was also surveyed and the subpopulation identity of these cultivars was determined using a set of genome-wide SSR and SNP markers (Garris et al., 2005; Caicedo et al., 2007). Overall, the *badh2.1* allele was detected in 17 (10%) of these accessions, with the fragrant allele detected at the highest frequency in *Group V* and at the lowest frequency in *temperate japonica* and *aus* (Table 4.1).

Table 4.1: Frequency of *badh2.1* allele in wild and cultivated rice

	# Individuals		
	Total	***badh2.1* allele**	**Frequency *badh2.1* allele (%)**
O. rufipogon/ *O. nivara*	280	0.5	0.2
O. sativa	176	17	10.0
indica	54	3	6.0
aus	23	0	0
Group V (aromatic)	10	6	60.0
temperate japonica	36	0	0
tropical japonica	53	8	15.0

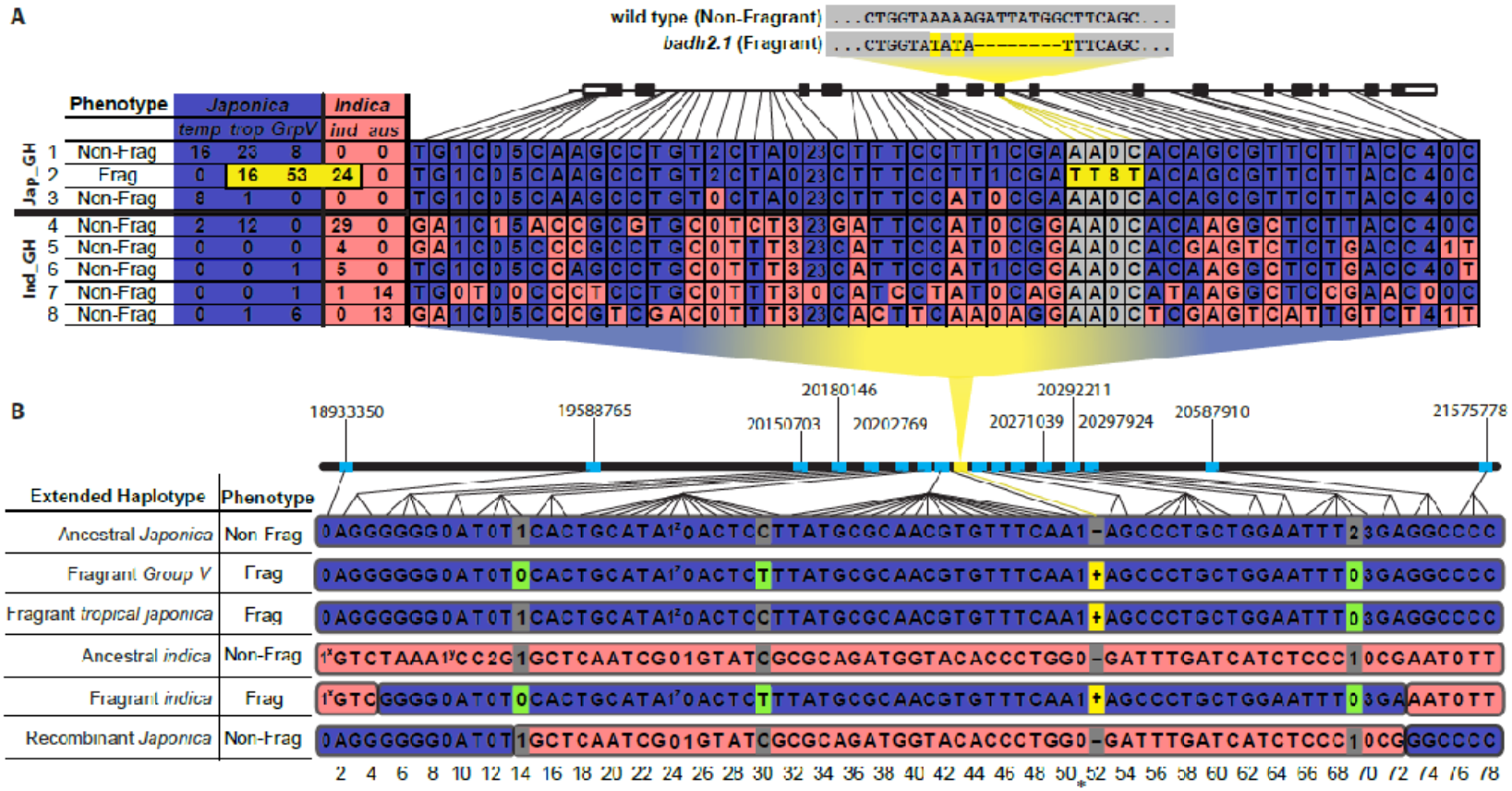
Origin of the *badh2.1* allele

Given that an identical *badh2.1* allele was detected in both the *Japonica* and *Indica* varietal groups, it was our goal to determine in which group it had originated. To address this question, we sequenced across the *BADH2* gene in a panel of 242 *O. sativa* accessions, which included the original panel and additional accessions known to possess the *badh2.1* allele (Fitzgerald et al., 2008) (Supplemental Table 4.1). In ~5 kilobases (kb) of aligned sequence, we detected 106 SNP, insertion-deletion (indel) and SSR polymorphisms, 54 of which were present at a frequency >5%. These polymorphisms were used to construct eight gene haplotypes (GH), and these haplotypes were clearly differentiated into two distinct clusters (Figure 4.2*A*; Supplemental Table 4.2). Within the first cluster, all accessions carrying the wild-type allele were from the *Japonica* varietal group (Jap_GH), while the majority of accessions from the second cluster (74%) were from the *Indica* varietal group (Ind_GH). Every accession carrying the *badh2.1* allele fell within the Jap_GH cluster, regardless of subpopulation identity. Within the Jap_GH cluster, fragrant accessions differed from the ancestral group of non-fragrant accessions only at the FNP that defines the *badh2.1* allele (highlighted in yellow in Figure 4.2*A*). These data therefore support a single origin of the *badh2.1* allele in a *Japonica*-like genetic background.

Figure 4.2: Haplotype Analysis of the *BADH2* Gene Region

4.2A: *BADH2* Gene Haplotypes. Sequence reads across the *BADH2* gene were aligned for 242 *O. sativa* accessions and all polymorphisms (frequency >5%) were concatenated and used to create eight gene haplotypes. Letters in each haplotype represent alternative nucleotides at a SNP site; numbers indicate the size of a deletion (0 = no deletion), with the relative position of each polymorphism indicated along the *BADH2* gene model. Haplotypes are numbered 1 though 8 followed by the corresponding fragrance phenotype, and the number of accessions from each subpopulation possessing that haplotype. Two gene haplotype clusters were identified: the *Japonica* Gene Haplotype Cluster (Jap_GH) and *Indica* Gene Haplotype Cluster (Ind_GH). Blue cells represent polymorphisms characteristic of Jap_GH, while red cells represent those characteristic of Ind_GH. The *badh2.1* FNP is depicted in gray/yellow, with yellow representing the fragrant allele.

4.2B: Extended Haplotypes. Out of 17 total extended haplotypes for *BADH2*, six "consensus" haplotypes are depicted with the phenotype of each indicated. Letter/number designations and color coding are as in Figure 4.2*A* (with the following abbreviated deletions: 1^{x}= 28, 1^{y}= 12, 1^{z}= 48). Breaks in coloration indicate positions where recombination was detected. Position 51, marked with a star, represents the MITE polymorphism described previously (Bourgis et al., 2008) . The *badh2.1* FNP is highlighted in yellow at site 52, with "+" referring to the derived (fragrant) allele and "–" referring to the wild-type (non-fragrant) allele. The polymorphisms highlighted in green at positions 14, 30, and 69 are those that were fixed within all fragrant *Group V* and *indica* accessions carrying the *badh2.1* allele. The relative positions and locations (on the TIGR v.5 pseudomolecule) of the 16 markers that provided AIPs are depicted as light blue dots along a black bar, representing the stretch of chromosome 8 that was sample-sequenced.


A
wild type (Non-Fragrant) ...CTGGTAAAAAGATTATGGCTTCAGC...
badh2.1 (Fragrant) ...CTGGTATATA--------TTTCAGC...
Phenotype
Japonica
Indica
temp trop GrpV
ind aus
Jap_GH
Ind_GH
1 Non-Frag 16 23 8 0 0
2 Frag 0 16 53 24 0
3 Non-Frag 8 1 0 0 0
4 Non-Frag 2 12 0 29 0
5 Non-Frag 0 0 0 4 0
6 Non-Frag 0 0 1 5 0
7 Non-Frag 0 0 1 1 14
8 Non-Frag 0 1 6 0 13
B
18933350
19588765
20150703
20202769
20180146
20271039
20292211
20297924
20587910
21575778
Extended Haplotype
Phenotype
Ancestral Japonica
Non Frag
Fragrant Group V
Frag
Fragrant tropical japonica
Frag
Ancestral indica
Non-Frag
Fragrant indica
Frag
Recombinant Japonica
Non-Frag
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78

All fragrant varieties from the *Indica* varietal group that carried the *badh2.1* allele clustered with Jap_GH, creating an apparent contradiction between the gene and genome phylogenies of these accessions. This incongruence can be explained if the fragrant *indica* accessions are found to carry a defined region of *Japonica*-like DNA around the *BADH2* gene within an *Indica* genetic background. We therefore broadened the scope of our haplotype analysis by sequencing 24 amplicons in a genomic region spanning 3.2 megabases (Mb) upstream and 2.1 Mb downstream of *BADH2* in our panel of 242 *O. sativa* accessions. In this 5.3 Mb region flanking *BADH2*, a total of 426 SNP, indel, and SSR polymorphisms were identified across ~13 kb of aligned sequence, with 271 polymorphisms present at a frequency >5%. From these, 78 ancestrally informative polymorphisms (AIPs) were identified and used to create extended haplotypes (see Materials and Methods; Supplemental Table 4.3). These extended haplotypes were then summarized into "consensus" extended haplotypes representing six major haplotype classes (Figure 4.2*B*). The extended haplotypes were consistent with the gene haplotypes, with all *badh2.1*-containing accessions having an extended region of *Japonica* DNA around the *BADH2* gene. In the 24 *indica* accessions carrying the *badh2.1* allele, the *Japonica* region was bordered by recombination breakpoints ~650 kb upstream and ~330 kb downstream, with the flanking regions identical to the ancestral *Indica* extended haplotype (Figure 4.2*B*). This supports the hypothesis that the *badh2.1* allele was transferred via introgression into *indica*.

Group V accessions carrying the *badh2.1* allele possessed an extended haplotype identical to ancestral *Japonica*, with the exception of the FNP and three unique polymorphisms at sites 14, 30, and 69 (highlighted in green; Figure 4.2*B*). These same three signature polymorphisms flanking the FNP were detected within the *Japonica*-

like region found in all *indica* accessions carrying the *badh2.1* allele. The single wild accession from Myanmar that was heterozygous for the *badh2.1* allele segregated 1:2:1 in the next generation, and haplotype analysis confirmed the chromosome carrying the *badh2.1* allele also contained the same three *Group V*-specific polymorphisms described above. We were therefore able to trace the ancestry of the *badh2.1* allele in the heterozygous wild accession and all fragrant *indica* accessions to a *Group V* ancestor.

There were several non-fragrant *Japonica* accessions that grouped with the Ind_GH cluster, which was otherwise made up of non-fragrant *Indica* accessions (Figure 4.2*A*). These non-fragrant accessions from the *Japonica* varietal group contained an *Indica*-like genomic region around the *BADH2* gene, and all showed recombination back to ancestral *Japonica* in the flanking regions (Figure 4.2*B*; "Recombinant *Japonica*"). This provides an explanation for the number of *Japonica* accessions having a *BADH2* gene haplotype that clustered with Ind_GH.

Reduced nucleotide diversity and elevated linkage disequilibrium at the *badh2.1* allele

To examine evidence of selection around the *badh2.1* allele, we analyzed the nucleotide diversity across the *BADH2* gene in our panel of 242 accessions. We compared the total nucleotide diversity (θ_π) in the ~5 kb of aligned sequence across the *BADH2* gene between accessions carrying the wild type and *badh2.1* alleles. Fragrant accessions carrying the *badh2.1* allele exhibited a 97.4% reduction in diversity on average compared to the non-fragrant accessions (Table 4.2).

Table 4.2: Average Nucleotide Diversity across *BADH2* gene (θ_π per kb)						
	indica	*aus*	*Group V (aromatic)*	*temperate japonica*	*tropical japonica*	All
Badh2 (non-fragrant) (n=151)	3.79	5.25	6.75	2.13	4.50	6.64
badh2.1 (fragrant) (n=93)	0.03	None	0.06	None	0.29	0.17

Extended haplotype homozygosity (EHH) estimates the probability that two randomly chosen genomic regions are identical by descent, and allows the measurement of linkage disequilibrium (LD) decay with increasing distance from a functional mutation (Sabeti et al., 2002). We calculated the EHH for the *BADH2* gene in our germplasm panel by comparing the extended haplotypes of the fragrant (*badh2.1*) and non-fragrant accessions. The fragrant accessions carrying the *badh2.1* allele exhibited a large block of extended LD around the FNP, while LD around the wild-type allele declined rapidly (Figure 4.3). This same pattern was observed when the extent of LD was compared between accessions containing the wild-type or the *badh2.1* allele within individual subpopulations (Supplemental Figure 4.1). The dramatically reduced nucleotide diversity at *badh2.1*, coupled with the extended region of linkage disequilibrium surrounding this allele, provide evidence for strong positive selection for the *badh2.1* FNP.

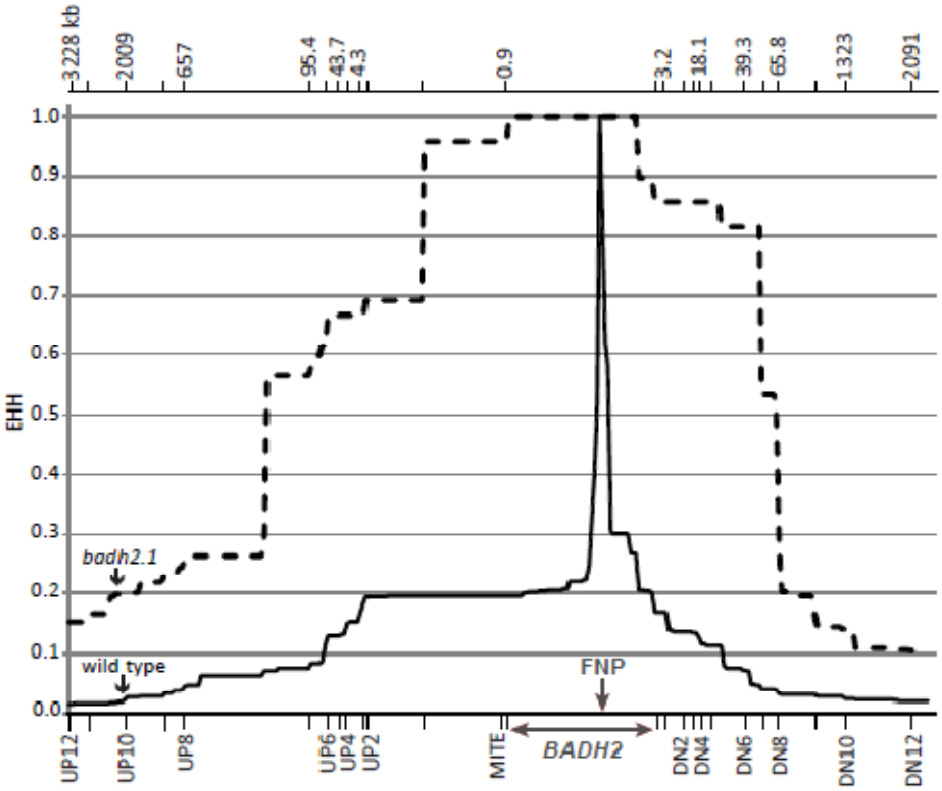


Figure 4.3: Extended Haplotype Homozygosity (EHH) across the *BADH2* genomic region. EHH values were calculated for the 242 *O. sativa* accessions examined in this study based on haplotype data across a 5.3 Mb genomic region surrounding the *BADH2* gene. Solid and dashed lines indicate the combined EHH values of accessions having the wild type and *badh2.1* alleles, respectively. The position of the *BADH2* gene is indicated and *badh2.1* FNP is indicated with an arrow. Dashes along the X-axis represent the locations of each amplicon used to obtain haplotype data (Supplemental Table 4.5). The physical distances of selected amplicons from the *BADH2* gene (in kb) are depicted across the top.

Additional mutations in the *BADH2* gene

In order to detect additional mutations in *BADH2* that may be responsible for enhanced 2AP synthesis, the *BADH2* gene was sequenced in 26 fragrant accessions that lacked either of the previously described derived alleles conferring fragrance (Fitzgerald et al., 2008). Sequence alignments from these 26 accessions with the 242 previously sequenced accessions revealed eight new nonsynonymous polymorphisms, four of which were frameshift-inducing indels and one of which was a SNP creating a premature stop codon, all putatively resulting in a truncated BADH2 protein (Figure 4.4*A*; Supplemental Table 4.4). The other three potentially functional polymorphisms included a 3 bp insertion, and two different SNPs in the coding region. While several of these polymorphisms were found and confirmed in only a single accession, four were found in multiple fragrant accessions, and these polymorphisms appear to have strong geographic associations (Figure 4.4*B*). There were two accessions for which we do not, as yet, have candidate functional mutations that could explain the presence of elevated 2AP levels.

Figure 4.4: *BADH2* Allelic Diversity

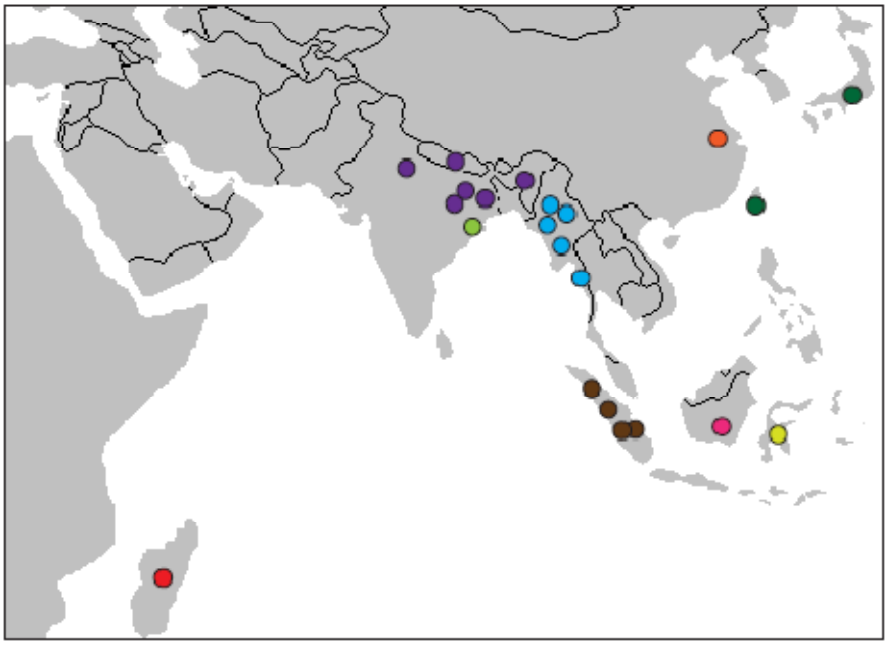
4.4A: Coding Mutations in the *BADH2* gene. Sequencing across 26 fragrant accessions that lacked either known mutation in *BADH2* identified eight novel mutations in the coding region of the gene that are putatively responsible for the fragrance phenotype. The gene model for *BADH2*, with 15 exons and 14 introns, is depicted for each allele along with the location and identity of the coding mutation. Four white stars in the gene model of the wild type allele illustrate the locations of critical catalytic or substrate-binding domains while the orange star indicates the location of an oligimerization domain (Chen et al., 2008). For each allele, the number (N) and genetic subpopulation(s) of accessions having that allele are shown, along with the average 2AP concentration of the accessions possessing that allele (with standard deviations). We obtained accurate 2AP measurements for only one of the accessions carrying the *badh2.10* allele, so no standard deviation is indicated.

4.4B: Geographic locations of accessions possessing novel mutations in *BADH2*. Each accession having a novel mutation in *BADH2* that is predicted to cause fragrance was placed on a map of Asia using all available collection/passport data for the accessions. The clustering of accessions with a given mutation within the same geographical region suggests that the fragrance trait was selected on multiple occasions, but these novel alleles remained isolated within local rice gene pools.

A

Allele	Mutation	N	Subpopulation	Avg [2AP]
wild type		149	ALL	< 0.05
badh2.1	FNP	93	*Group V, indica, tropical japonica*	0.70 ± 0.41
badh2.2	7bp deletion	1	*tropical japonica*	0.66
badh2.3	2bp deletion	1	*tropical japonica*	0.74
badh2.4	1 bp insertion	1	*tropical japonica*	0.59
badh2.5	1bp deletion	1	*indica*	0.70
badh2.6	G→T SNP	1	*aus*	0.41
badh2.7	1bp insertion	6	*aus*	0.43 ± 0.17
badh2.8	3bp insertion	5	*Group V*	0.36 ± 0.14
badh2.9	G→T SNP	6	*tropical japonica*	0.18 ± 0.08
badh2.10	C→T SNP	2	*tropical japonica*	0.35

B



DISCUSSION

A *Japonica* origin of fragrance in rice

This study presents an in-depth survey of the genetic diversity of the *BADH2* gene in a large panel of genetically and geographically diverse rice germplasm from across Asia. Fragrant accessions carrying the *badh2.1* allele exhibited a dramatic reduction in nucleotide diversity (97%) and elevated linkage disequilibrium around the gene compared to non-fragrant accessions, consistent with strong positive selection for the *badh2.1* allele. The intensity of selection on the *badh2.1* allele was similar to reports from other rice genes controlling grain morphology traits (i.e., *wx, rc, gs3*) (Olsen and Purugganan, 2002; Sweeney et al., 2007; Takano-Kai et al., 2009). Haplotype analysis allowed us to demonstrate that the *badh2.1* allele arose in the genetic background of the *Japonica* varietal group. The presence of the *badh2.1* allele within all Jasmine-like accessions, which are members of the *Indica* varietal group, prompted us to perform a more detailed analysis of the genomic region flanking the *badh2.1* allele. Extended haplotype analysis revealed a clear introgression of a *Japonica* genomic region encompassing the *badh2.1* allele in all fragrant *Indica* accessions. Additionally, there were three polymorphisms flanking the *badh2.1* allele that were diagnostic of the *Group V* subpopulation, which includes the highly valued Basmati accessions from India and Pakistan. These same three polymorphisms were found within all fragrant *Indica* accessions (ie, Jasmine), demonstrating for the first time that this important grain quality allele originated in the *Group V* subpopulation.

The derived polymorphism at site 69 that was ancestrally informative and fixed within fragrant *Group V* accessions was also fixed within all *tropical japonica* accessions carrying the *badh2.1* allele. This suggests that: 1) the *badh2.1* allele may have arisen in an ancestor common to both the *Group V* and *tropical japonica* lineages, and the

two *Group V*-specific polymorphisms (sites 14 and 30; Figure 4.2*B*) arose in the *Group V* lineage following the divergence of these groups or 2) that a genomic region containing the *badh2.1* allele and one downstream polymorphism (site 69) may have been introgressed from *Group V* to *tropical japonica*. The extended region of identical sequence flanking the *BADH2* gene between fragrant *Group V* and *tropical japonica* accessions precludes the detection of recombination breakpoints that would define an introgression from *Group V* to *tropical japonica* (or vice versa). [A more in-depth sequence analysis was used to develop several hypotheses for the origin of the *badh2.1* allele; presented in Chapter 8.]

Evidence from this study suggests that the *badh2.1* allele was selected as a *de novo* mutation in *O. sativa* after domestication from its wild progenitor, and presumably after the divergence of the *Japonica* subpopulations, given the high frequency of the allele in *Group V* and vanishingly low frequency in *temperate japonica*. We examined 280 diverse accessions of *O. rufipogon*/*O. nivara* from across Asia and found only a single accession from Myanmar that was heterozygous for the *badh2.1* allele. Sequence analysis of this accession demonstrated that the chromosome carrying the *badh2.1* allele contained the three signature polymorphisms fixed in all fragrant *Group V* accessions bordered by DNA sequence characteristic of *Indica*. While *badh2.1* has been identified in other wild rice germplasm from Southeast Asia (Vanavichit, 2004, 2007; Prathepha, 2009), it is likely that these represent introgressions of the derived allele into wild and weedy rice populations growing in the vicinity of fragrant cultivars (Bourgis et al., 2008). Gene flow of alleles from *O. sativa* into wild rice populations has been reported on multiple occasions, and is now thought to be ubiquitous (Song et al., 2003; Chen et al., 2004).

This study provides new understanding of the genetic relationships between *O. sativa* subpopulations and further clarifies the evolutionary history of Asian cultivated rice. Basmati varieties from the *Group V* subpopulation are often erroneously referred to as members of the *Indica* varietal group (Maqbool et al., 1998; Garg et al., 2002; Bashir et al., 2004). While varieties from the *indica* and *Group V* subpopulations are widely grown in South and Central Asia and both may exhibit long, slender grain morphology, researchers have long recognized the high levels of hybrid sterility in crosses between these two groups (Engle et al., 1969; Shobha, 1992; Singh et al., 2000; Khush and Dela Cruz, 2001). Glaszmann postulated an explanation for these hybrid incompatibilities by demonstrating with isozyme markers that *Group V* was clearly distinct from *Indica*, and that it grouped more closely with *Japonica* (Glaszmann, 1985, 1987). Additional examinations of the genetic diversity in *O. sativa* using chloroplast markers, nuclear SSRs, and SNPs independently determined that the *Group V* subpopulation is a unique genetic entity closely related to the *Japonica* varietal group (Jain et al., 2004; Garris et al., 2005; Caicedo et al., 2007). Our haplotype analysis in this study demonstrated that the *Group V* accessions, both fragrant and non-fragrant, cluster with the ancestral *Japonica* accessions both across the *BADH2* gene and across the entire 5.3 Mb region of chromosome 8 surveyed. This provides further evidence that *Group V*, despite its morphological similarities and geographical distribution that overlaps with *Indica*, is genetically a member of the *Japonica* varietal group. Interestingly, the overlapping distributions of *Group V* and *Indica* varieties in South Asia may have provided a corridor for the transfer of the major fragrance allele to *Indica*.

BADH2 **in the broader context of rice domestication and varietal differentiation**

The alleles associated with rice domestication and grain quality enhancement followed several distinct evolutionary paths (Kovach et al., 2007). Some of these alleles appear to have been selected from standing variation in the wild, where a beneficial mutation was harnessed very early during the domestication process. Other critical alleles would have been selected *de novo* from within a single gene pool, and these alleles either remained isolated or became widely disseminated among the subpopulations of *O. sativa*. Many of the derived alleles characterized thus far in rice are shared by both the *Indica* and *Japonica* varietal groups, with evidence of a single origin within one of these groups, consistent with the process of introgression. The *rc* allele, which is responsible for white pericarp in 98% of modern *O. sativa* varieties, was the first reported example of a domestication-related gene that arose in a *Japonica* genetic background followed by dissemination into other, genetically divergent subpopulations (Sweeney et al., 2007). The average size of the *rc*-containing introgression in the *Indica* varietal group was around 1 Mb, similar to the size of the *badh2.1* introgression in *indica* as reported in this study. Similarly, a mutation in the *GS3* gene conferring long grain originated in a *Japonica* ancestor, with subsequent introgression into long-grained *indica* cultivars (Takano-Kai et al., 2009). The evolutionary history of the *Waxy* gene responsible for glutinous or "sticky" rice followed a similar trajectory, with 97% of low amylose accessions having the glutinous allele (*wx*) originating from the *Japonica* varietal group, Wx^b (Izawa, 2008). In this study, we demonstrate that the evolutionary history of the *badh2.1* allele is also consistent with this pattern. These alleles would have been transferred between the rice subpopulations through the process of introgressive hybridization, facilitated by the higher outcrossing rates among early rice cultivars and the physical proximity of divergent (*Indica* and *Japonica*) cultivars as a result of population expansion and

human migration throughout Asia (Kovach et al., 2007; Sang and Ge, 2007; Vaughan et al., 2008; Izawa et al., 2009). It is noteworthy that despite the apparent importance of hybridization and gene flow during rice evolution, opposing forces have maintained the genetic divergence among the subpopulations of *O. sativa*. Solving this paradox will require future research to identify the key factors that contribute to subpopulation isolation, as well as to provide insight into the dynamics of genetic exchange among these groups.

Independent origins of fragrance in rice

The presence of rice varieties exhibiting elevated 2AP levels, but lacking any known nonfunctional allele of *BADH2*, raised the possibility that there might be additional fragrance-causing alleles of *BADH2* (Fitzgerald et al., 2008). Analysis of 26 fragrant accessions that lacked any known fragrance allele identified eight new polymorphisms in the coding region of the gene that are predicted to alter the BADH2 protein. It had been previously shown that only the full-length *BADH2* transcript, resulting in an intact 503 amino acid protein, was capable of inhibiting 2AP production (Chen et al., 2008). Four of the eight new coding polymorphisms identified in this study (creating alleles *badh2.3-2.6*) are predicted to cause premature transcript termination, which would putatively abolish protein function and result in fragrance. These mutations all result in truncation of the *BADH2* protein prior to critical residues that form the catalytic and/or substrate binding domains. Additionally, the *badh2.7* allele, which was shared by six *aus* accessions, is also predicted to result in a shortened transcript that would eliminate an oligomerization domain of the protein. The other three new alleles (*badh2.8-2.10*) either result in an additional in-frame amino acid (*badh2.8*) or amino acid substitutions (*badh2.9-2.10*). Despite the association of fragrance with

these coding mutations in the *BADH2* gene, future work is needed to confirm the effect of these mutations on the degree of 2AP accumulation.

The geographic association between accessions carrying the same mutant alleles at *BADH2* suggests fragrance was selected independently on multiple occasions in different geographic regions. Interestingly, while we now know the predominant allele responsible for enhanced 2AP production in rice originated within the *Japonica* varietal group, it appears that other fragrant alleles of *BADH2* were identified within the *Indica* varietal group, exemplified by the *badh2.7* allele, which was found in several *aus* accessions. These and other findings suggest that the *Group V* and *aus* lineages harbor useful alleles that have yet to be fully exploited in rice improvement (Garris et al., 2003; Garris et al., 2005; Xu et al., 2006; Sweeney et al., 2007).

Out of the 26 accessions exhibiting fragrance but lacking any of the previously identified fragrance alleles, there were two accessions for which we could find no mutation in the coding or promoter regions that would be predicted to alter the BADH2 protein or its expression. It is therefore possible that these accessions contain genetic lesions elsewhere in the metabolic pathway controlling 2AP synthesis; mapping experiments are underway to test this hypothesis. Identifying additional genes responsible for fragrance would provide new opportunities to tailor rice grain quality to suit cultural preferences (Fitzgerald et al., 2009).

APPENDIX

Supplemental Table 4.1: Rice Accessions Used in This Study

Wild Accessions:					
NSFTV #	Gene Bank #	Cornell RA #	Country of Origin	Species	*BADH2* allele
401_C	IRGC 80433	RA 6970	India	O. rufipogon	WT
402_B	IRGC 80539	RA 6971	India	O. spontanea	WT
403_A	IRGC 80562	RA 6972	India	O. rufipogon	WT
404_A	IRGC 80582	RA 6973	India	O. nivara	WT
405_A	IRGC 80586	RA 6974	India	O. spontanea	WT
406_A	IRGC 80592	RA 6975	India	O. rufipogon	WT
407_C	IRGC 80742	RA 6976	Myanmar	O. rufipogon	WT
408_B	IRGC 80745	RA 6977	Myanmar	O. spontanea	WT
410_A	IRGC 80759	RA 6979	Myanmar	O. nivara	WT
411_B	IRGC 81801	RA 6980	Indonesia	O. rufipogon	WT
412_C	IRGC 81802	RA 6981	Indonesia	O. rufipogon	WT
413_A	IRGC 81850	RA 6982	India	O. nivara	WT
414_C	IRGC 81903	RA 6983	India	O. spontanea	WT
415_B	IRGC 81909	RA 6984	India	O. spontanea	WT
416_A	IRGC 81970	RA 6985	Thailand	O. spontanea	WT
417_B	IRGC 81976	RA 6986	Indonesia	O. rufipogon	WT
418_A	IRGC 81977	RA 6987	Indonesia	O. rufipogon	WT
420_A	IRGC 81984	RA 6989	Laos	O. rufipogon	WT
421_C	IRGC 81990	RA 6990	Myanmar	O. rufipogon	WT
422_A	IRGC 81993	RA 6991	Vietnam	O. rufipogon	WT
423_A	IRGC 81994	RA 6992	Papua New Guinea	O. rufipogon	WT
424_C	IRGC 81996	RA 6993	Papua New Guinea	O. rufipogon	WT
425_B	IRGC 82040	RA 6994	Thailand	O. rufipogon	WT
426_C	IRGC 82979	RA 6995	Thailand	O. rufipogon	WT
427_B	IRGC 82988	RA 6996	China	O. rufipogon	WT
428_A	IRGC 82989	RA 6997	China	O. rufipogon	WT
429_A	IRGC 82990	RA 6998	China	O. rufipogon	WT
430_B	IRGC 82991	RA 6999	China	O. rufipogon	WT
431_A	IRGC 82992	RA 7000	China	O. rufipogon	WT
432_B	IRGC 83794	RA 7001	Thailand	O. rufipogon	WT
433_A	IRGC 83795	RA 7002	India	O. rufipogon	WT
434_A	IRGC 83823	RA 7003	Vietnam	O. rufipogon	WT
435_C	IRGC 86448	RA 7004	Thailand	O. rufipogon	WT
436_A	IRGC 86454	RA 7005	Vietnam	O. rufipogon	WT
437_B	IRGC 86475	RA 7006	India	O. rufipogon	WT
438_B	IRGC 86476	RA 7007	India	O. rufipogon	WT
439_A	IRGC 86486	RA 7008	Thailand	O. rufipogon	WT
440_B	IRGC 88787	RA 7009	Bangladesh	O. rufipogon	WT
441_A	IRGC 92605	RA 7010	Indonesia	O. rufipogon	WT
442_A	IRGC 93181	RA 7011	Nepal	O. nivara	WT
443_B	IRGC 93183	RA 7012	Nepal	O. nivara	WT
444_A	IRGC 93188	RA 7013	Nepal	O. nivara	WT
445_B	IRGC 93189	RA 7014	Nepal	O. nivara	WT
446_A	IRGC 93224	RA 7015	Nepal	O. spontanea	WT
447_C	IRGC 93274	RA 7016	Indonesia	O. rufipogon	WT
449_A	IRGC 100195	RA 7018	Myanmar	O. nivara	WT
450_A	IRGC 100916	RA 7019	China	O. rufipogon	WT
451_B	IRGC 101508	RA 7020	India	O. nivara	WT
452_A	IRGC 103308	RA 7021	Taiwan	O. rufipogon	WT
453_C	IRGC 103404	RA 7022	Bangladesh	O. rufipogon	WT
454_B	IRGC 103821	RA 7023	China	O. nivara	WT
455_A	IRGC 103823	RA 7024	China	O. rufipogon	WT
456_A	IRGC 103824	RA 7025	China	O. nivara	WT
457_B	IRGC 103838	RA 7026	Bangladesh	O. nivara	WT
458_A	IRGC 103844	RA 7027	Bangladesh	O. rufipogon	WT
459_B	IRGC 103847	RA 7028	India	O. rufipogon	WT
460_C	IRGC 103848	RA 7029	India	O. rufipogon	WT
461_A	IRGC 104057	RA 7030	China	O. rufipogon	WT
462_A	IRGC 104501	RA 7031	India	O. rufipogon	WT
463_B	IRGC 104599	RA 7032	Sri Lanka	O. rufipogon	WT
464_A	IRGC 104602	RA 7033	Sri Lanka	O. rufipogon	WT
465_B	IRGC 104620	RA 7034	China	O. spontanea	WT
467_B	IRGC 104624	RA 7036	China	O. rufipogon	WT
468_B	IRGC 104626	RA 7037	China	O. spontanea	WT
469_A	IRGC 104628	RA 7038	China	O. spontanea	WT
470_A	IRGC 104632	RA 7039	China	O. spontanea	WT
471_C	IRGC 104634	RA 7040	China	O. spontanea	WT
472_B	IRGC 104636	RA 7041	China	O. spontanea	WT
473_C	IRGC 104644	RA 7042	Thailand	O. nivara	WT
474_A	IRGC 104714	RA 7043	Thailand	O. rufipogon	WT
475_B	IRGC 104823	RA 7044	Thailand	O. nivara	WT
476_A	IRGC 104959	RA 7045	China	O. spontanea	WT
477_A	IRGC 104967	RA 7046	China	O. spontanea	WT
478_B	IRGC 104971	RA 7047	China	O. spontanea	WT
479_A	IRGC 105220	RA 7048	Indonesia	O. officianalis	WT

Supplemental Table 4.1 (continued)

480_B	IRGC 105250	RA 7049	Thailand	O. rufipogon	WT
481_C	IRGC 105343	RA 7050	India	O. nivara	WT
482_B	IRGC 105349	RA 7051	India	O. rufipogon	WT
483_C	IRGC 105375	RA 7052	Thailand	O. rufipogon	WT
484_C	IRGC 105388	RA 7053	Thailand	O. rufipogon	WT
485_A	IRGC 105400	RA 7054	China	O. rufipogon	WT
486_C	IRGC 105402	RA 7055	China	O. rufipogon	WT
487_C	IRGC 105428	RA 7056	Sri Lanka	O. nivara	WT
488_B	IRGC 105491	RA 7057	Malaysia	O. rufipogon	WT
489_A	IRGC 105564	RA 7058	Indonesia	O. spontanea	WT
490_C	IRGC 105567	RA 7059	Indonesia	O. rufipogon	WT
491_B	IRGC 105568	RA 7060	Philippines	O. rufipogon	WT
492_B	IRGC 105569	RA 7061	Cambodia	O. rufipogon	WT
493_A	IRGC 105706	RA 7062	Nepal	O. nivara	WT
494_A	IRGC 105711	RA 7063	India	O. rufipogon	WT
495_A	IRGC 105717	RA 7064	Cambodia	O. nivara	WT
496_A	IRGC 105720	RA 7065	Cambodia	O. rufipogon	WT
498_A	IRGC 105735	RA 7067	Cambodia	O. rufipogon	WT
499_B	IRGC 105767	RA 7068	Thailand	O. rufipogon	WT
500_A	IRGC 105785	RA 7069	Thailand	O. nivara	WT
501_B	IRGC 105821	RA 7070	Thailand	O. nivara	WT
503_B	IRGC 105843	RA 7072	Thailand	O. rufipogon	WT
504_A	IRGC 105847	RA 7073	Thailand	O. rufipogon	WT
505_A	IRGC 105855	RA 7074	Thailand	O. rufipogon	WT
507_C	IRGC 105881	RA 7076	Bangladesh	O. rufipogon	WT
508_A	IRGC 105890	RA 7077	Bangladesh	O. rufipogon	WT
509_A	IRGC 105897	RA 7078	Bangladesh	O. rufipogon	WT
510_A	IRGC 105898	RA 7079	Bangladesh	O. rufipogon	WT
511_B	IRGC 105909	RA 7080	Thailand	O. rufipogon	WT
512_A	IRGC 105942	RA 7081	Thailand	O. rufipogon	WT
513_A		RA 7082	Indonesia	O. rufipogon	WT
514_C	IRGC 105956	RA 7083	Indonesia	O. rufipogon	WT
515_A	IRGC 105958	RA 7084	Indonesia	O. rufipogon	WT
516_B	IRGC 106036	RA 7085	Malaysia	O. rufipogon	WT
517_A	IRGC 106057	RA 7086	India	O. rufipogon	WT
518_A	IRGC 106078	RA 7087	India	O. rufipogon	WT
519_B	IRGC 106115	RA 7088	India	O. rufipogon	WT
520_C		RA 7089	India	O. rufipogon	WT
521_C	IRGC 106145	RA 7090	Laos	O. rufipogon	WT
522_B	IRGC 106150	RA 7091	Laos	O. rufipogon	WT
523_A	IRGC 106155	RA 7092	Laos	O. nivara	WT
524_C	IRGC 106156	RA 7093	Laos	O. rufipogon	WT
525_B	IRGC 106161	RA 7094	Laos	O. rufipogon	WT
526_A		RA 7095	Laos	O. rufipogon	WT
528_A	IRGC 106168	RA 7097	Vietnam	O. rufipogon	WT
529_A	IRGC 106169	RA 7098	Vietnam	O. rufipogon	WT
530_B		RA 7099	Papua New Guinea	O. rufipogon	WT
531_C		RA 7100	Papua New Guinea	O. rufipogon	WT
533_B	IRGC 106327	RA 7102	Cambodia	O. rufipogon	WT
534_C	IRGC 106332	RA 7103	Cambodia	O. rufipogon	WT
535_C	IRGC 106342	RA 7104	Myanmar	O. rufipogon	WT
536_A	IRGC 106357	RA 7105	Myanmar	O. rufipogon	WT
537_B	IRGC 106384		Myanmar	O. spontanea	WT
538_A	IRGC 106410	RA 7107	Vietnam	O. rufipogon	WT
539_A		RA 7108	Vietnam	O. rufipogon	WT
540_C	IRGC 106413	RA 7109	Vietnam	O. rufipogon	WT
541_B	IRGC 106414	RA 7110	Vietnam	O. rufipogon	WT
542_A	IRGC 106415		Vietnam	O. rufipogon	WT
543_A	IRGC 106420	RA 7112	Vietnam	O. rufipogon	WT
544_B	IRGC 106452		Indonesia	O. rufipogon	WT
545_A	IRGC 106453	RA 7114	Indonesia	O. rufipogon	WT
546_B	IRGC 106509	RA 7115	Myanmar	O. rufipogon	WT
547_A	IRGC 105908	RA 7116	Thailand	O. rufipogon	WT
549_C	IRGC 81881	RA 7118	India	O. rufipogon	WT
550_B	IRGC 81887	RA 7119	India	O. rufipogon	WT
551_B	IRGC 100596	RA 6768	Taiwan	O. rufipogon/O. nivara	WT
552_C	IRGC 100920	RA 6770	Malaysia	O. rufipogon x O. nivara	WT
553_B	IRGC 100926	RA 6771	Myanmar	O. rufipogon	Heterozygous (WT/*badh2.1*)
554_A	IRGC 103305	RA 6772	Philippines	O. rufipogon	WT
554_B	IRGC 103305		Philippines	O. rufipogon	WT
555_B	IRGC 105349	RA 6773	India	O. rufipogon	WT
556_A	IRGC 105494	RA 6775	Myanmar	O. rufipogon	WT
557_B	IRGC 105567	RA 6776	Indonesia	O. rufipogon	WT
558_A	IRGC 105616	RA 6777	unknown	O. rufipogon	WT
559_C	IRGC 105618	RA 6778	unknown	O. rufipogon	WT
560_B		RA 6779	Cambodia	O. rufipogon	WT
561_C	IRGC 105868	RA 6780	Bangladesh	O. rufipogon	WT
562_C		RA 6781	Bangladesh	O. rufipogon	WT
563_A		RA 6782	Indonesia	O. rufipogon	WT
565_B		RA 6784	India	O. rufipogon	WT

Supplemental Table 4.1 (continued)

566_B		RA 6785	Laos	O. rufipogon	WT
567_C	IRGC 106167	RA 6786	Vietnam	O. rufipogon	WT
568_A	IRGC 106263	RA 6787	Papua New Guinea	O. rufipogon	WT
569_A	IRGC 106264	RA 6788	Papua New Guinea	O. rufipogon	WT
570_A	IRGC 106266	RA 6789	Papua New Guinea	O. rufipogon	WT
574_A	IRGC 106270		Papua New Guinea	O. rufipogon	WT
574_B	IRGC 106270	RA 6793	Papua New Guinea	O. rufipogon	WT
575_A	IRGC 106272	RA 6794	Papua New Guinea	O. rufipogon	WT
576_C		RA 6795	Papua New Guinea	O. rufipogon	WT
577_A	IRGC 106274	RA 6796	Papua New Guinea	O. rufipogon	WT
578_B	IRGC 106275	RA 6797	Papua New Guinea	O. rufipogon	WT
579_C	IRGC 106276	RA 6798	Papua New Guinea	O. rufipogon	WT
581_A	IRGC 106278	RA 6800	Papua New Guinea	O. rufipogon	WT
582_C	IRGC 106279	RA 6801	Papua New Guinea	O. rufipogon	WT
583_B	IRGC 106280	RA 6802	Papua New Guinea	O. rufipogon	WT
584_A	IRGC 106282	RA 6803	Papua New Guinea	O. rufipogon	WT
585_B		RA 6804	Papua New Guinea	O. rufipogon	WT
587_B	IRGC 106285	RA 6806	Papua New Guinea	O. rufipogon	WT
588_C	IRGC 106286	RA 6807	Papua New Guinea	O. rufipogon	WT
590_C	IRGC 106289	RA 6809	Papua New Guinea	O. rufipogon	WT
591_A	IRGC 106290	RA 6810	Papua New Guinea	O. rufipogon	WT
592_B	IRGC 80671	RA 6811	India	O. rufipogon	WT
593_A	IRGC 105757	RA 6812	Thailand	O. rufipogon	WT
594_C		RA 6813	Vietnam	O. rufipogon	WT
599_A	IRGC 100183	RA 2772	India	O. rufipogon x O. sativa	WT
600_B	IRGC 100187	RA 2773	Malaysia	O. sativa/O. rufipogon	WT
602_A	IRGC 100900	RA 2775	India	O. nivara/O. rufipogon	WT
604_C	IRGC 100907	RA 2777	Taiwan	O. rufipogon/O. nivara	WT
605_C	IRGC 100911	RA 2778	Thailand	O. nivara/O. rufipogon	WT
655_B	IRGC 100898	RA 7757	India	O. nivara	WT
656_B	IRGC 104443	RA 7758	Thailand	O. nivara	WT
657_A	IRGC 104705	RA 7759	India	O. nivara	WT
658_B	IRGC 100912	RA 7760	Thailand	O. nivara/O. rufipogon	WT
659_B	IRGC 80774	RA7851	Philippines	O. rufipogon	WT
660_B	IRGC 82993	RA7852	China	O. rufipogon	WT
661_C	IRGC 99554	RA7853	China	O. rufipogon	WT
662_C	IRGC 99555	RA7854	China	O. rufipogon	WT
663_C	IRGC 99556	RA7855	China	O. rufipogon	WT
664_A	IRGC 100196	RA7856	Myanmar	O. nivara	WT
665_A	IRGC 100203	RA7857	Myanmar	O. rufipogon/O. sativa	WT
666_B	IRGC 100211	RA7858	India	O. rufipogon	WT
668_B	IRGC 100588	RA7860	Taiwan	O. rufipogon	WT
669_C	IRGC 100593	RA7861	Taiwan	O. nivara	WT
670_C	IRGC 100597	RA7862	Taiwan	O. rufipogon	WT
671_A	IRGC 100599	RA7863	Taiwan	O. rufipogon	WT
672_B	IRGC 100639	RA7864	Taiwan	O. rufipogon	WT
673_B	IRGC 100647	RA7865	Taiwan	O. rufipogon	WT
674_C	IRGC 100657	RA7866	Taiwan	O. rufipogon	WT
675_C	IRGC 100678	RA7867	Taiwan	O. rufipogon	WT
676_C	IRGC 100692	RA7868	Taiwan	O. rufipogon	WT
677_A	IRGC 100897	RA7869	India	O. nivara	WT
678_C	IRGC 100898	RA7870	India	O. nivara	WT
679_B	IRGC 100899	RA7871	India	O. nivara	WT
680_B	IRGC 100902	RA7872	India	O. nivara/O. rufipogon	WT
681_C	IRGC 100903	RA7873	India	O. nivara	WT
682_A	IRGC 100904	RA7874	Thailand	O. rufipogon	WT
683_A	IRGC 100918	RA7875	Cambodia	O. nivara	WT
684_A	IRGC 100920	RA7876	Malaysia	O. rufipogon/O. nivara	WT
685_A	IRGC 100923	RA7877	Myanmar	O. rufipogon	WT
686_C	IRGC 100926	RA7878	Myanmar	O. nivara	WT
687_A	IRGC 101193	RA7879	Taiwan	O. rufipogon/O. sativativa	WT
688_A	IRGC 101450	RA7880	Taiwan	O. rufipogon	WT
690_C	IRGC 101942	RA7882	Malaysia	O. rufipogon	WT
691_A	IRGC 101967	RA7883	India	O. nivara	WT
692_C	IRGC 101979	RA7884	India	O. rufipogon	WT
693_A	IRGC 102116	RA7885	Cambodia	O. rufipogon	WT
694_B	IRGC 103306	RA7886	Taiwan	O. rufipogon	WT
695_A	IRGC 103407	RA7887	Sri Lanka	O. rufipogon	WT
696_B	IRGC 103415	RA7888	Sri Lanka	O. rufipogon	WT
698_A	IRGC 103418	RA7890	Sri Lanka	O. rufipogon	WT
700_A	IRGC 103423	RA7892	Sri Lanka	O. nivara	WT
701_B	IRGC 103813	RA7893	China	O. nivara	WT
702_B	IRGC 103814	RA7894	China	O. nivara/O. rufipogon	WT
703_A	IRGC 103817	RA7895	China	O. nivara	WT
704_B	IRGC 103818	RA7896	China	O. rufipogon	WT
705_B	IRGC 103822	RA7897	China	O. nivara	WT
706_B	IRGC 103825	RA7898	China	O. rufipogon/O. nivara	WT
707_B	IRGC 103835	RA7899	Bangladesh	O. nivara	WT
708_A	IRGC 103836	RA7900	Bangladesh	O. nivara	WT
709_A	IRGC 103837	RA7901	Bangladesh	O. nivara	WT

Supplemental Table 4.1 (continued)

710_B	IRGC 103840	RA7902	Bangladesh	O. nivara	WT		
711_A	IRGC 103841	RA7903	Bangladesh	O. nivara	WT		
712_B	IRGC 103845	RA7904	India	O. nivara	WT		
713_A	IRGC 104056	RA7905	China	O. rufipogon/O. nivara	WT		
714_C	IRGC 104058	RA7906	China	O. rufipogon/O. nivara	WT		
715_B	IRGC 104497	RA7907	Thailand	O. rufipogon/O. nivara	WT		
716_B	IRGC 104647	RA7908	Thailand	O. rufipogon	WT		
717_B	IRGC 104650	RA7909	Thailand	O. nivara	WT		
718_C	IRGC 104670	RA7910	Thailand	O. nivara	WT		
719_A	IRGC 104687	RA7911	India	O. nivara	WT		
720_A	IRGC 104703	RA7912	India	O. nivara	WT		
721_C	IRGC 104705	RA7913	India	O. nivara	WT		
722_A	IRGC 104962	RA7914	China	O. rufipogon/O. nivara	WT		
723_B	IRGC 104969	RA7915	China	O. rufipogon/O. nivara	WT		
724_C	IRGC 104999	RA7916	Indonesia	O. rufipogon	WT		
725_A	IRGC 105319	RA7917	India	O. nivara	WT		
727_B	IRGC 105391	RA7919	Thailand	O. nivara	WT		
728_A	IRGC 105397	RA7920	China	O. rufipogon/O. nivara	WT		
729_C	IRGC 105403	RA7921	China	O. rufipogon/O. nivara	WT		
731_C	IRGC 105424	RA7923	Sri Lanka	O. rufipogon	WT		
732_C	IRGC 105431	RA7924	Sri Lanka	O. nivara	WT		
733_B	IRGC 105444	RA7925	Sri Lanka	O. nivara	WT		
734_C	IRGC 105487	RA7926	Thailand	O. rufipogon	WT		
735_C	IRGC 105493	RA7927	Myanmar	O. rufipogon	WT		
737_B	IRGC 105599	RA7929	Thailand	O. rufipogon	WT		
738_B	IRGC 105601	RA7930	Thailand	O. rufipogon/O. nivara	WT		
740_C	IRGC 105622	RA7932	India	O. rufipogon	WT		
741_A	IRGC 105624	RA7933	India	O. nivara	WT		
742_C	IRGC 105625	RA7934	India	O. nivara	WT		
743_C	IRGC 105705	RA7935	NPL	O. nivara	WT		
744_A	IRGC 105716	RA7936	Cambodia	O. nivara	WT		
745_A	IRGC 105738	RA7937	Cambodia	O. rufipogon	WT		
746_B	IRGC 105740	RA7938	Cambodia	O. nivara	WT		
747_C	IRGC 105742	RA7939	Cambodia	O. nivara	WT		
748_C	IRGC 105763	RA7940	Thailand	O. nivara	WT		
749_C	IRGC 105867	RA7941	Thailand	O. nivara	WT		
751_C	IRGC 105895	RA7943	Bangladesh	O. nivara	WT		
752_A	IRGC 105901	RA7944	Bangladesh	O. rufipogon	WT		
753_B	IRGC 105960	RA7945	Bangladesh	O. rufipogon	WT		
755_B	IRGC 106103	RA7947	India	O. rufipogon	WT		
757_A	IRGC 106148	RA7949	Laos	O. nivara	WT		
759_A	IRGC 106336	RA7951	Cambodia	O. rufipogon	WT		
760_A	IRGC 106345	RA7952	Myanmar	O. nivara	WT		
762_C	IRGC 106396	RA7954	Myanmar	O. nivara	WT		

***O. sativa* accessions:**

NSFTV #	Gene Bank #	Cornell RA #	Accession Name	Country of Origin	Subpopulation	*BADH2* allele	[2AP]
001	IRGC 3135	RA 7137	Agostano	Italy	*temperate japonica*	WT	<0.05
003	IRGC 51250	RA 7139	Ai-Chiao-Hong	China	*indica*	WT	<0.05
004	IRGC 12386	RA 7140	ARC 10177	India	*aus*	WT	<0.05
005	IRGC 12440	RA 7141	ARC 10352	India	*Group V (aromatic)*	*badh2.1*	0.48
006	IRGC 12331	RA 7142	ARC 7229	India	*aus*	WT	N/A
007	IRGC 43325	RA 7143	Arias	Indonesia	*tropical japonica*	*badh2.1*	1.07
008	IRGC 6949	RA 7144	Asse Y Pung	Philippines	*tropical japonica*	*badh2.1*	0.42
009	IRGC 33984	RA 7145	Baber	India	*tropical japonica*	WT	N/A
011	IRGC 6590	RA 7147	Baugamon 14	Bangladesh	*aus*	WT	<0.05
012	IRGC 27805	RA 7148	Basmati	Pakistan	*Group V (aromatic)*	WT	<0.05
013	IRGC 27798	RA 7149	Basmati 1	Pakistan	*aus*	WT	<0.05
014	IRGC 53637	RA 7150	Basmati 217	India	*Group V (aromatic)*	*badh2.1*	N/A
015	IRGC 55457	RA 7151	Beonjo	Korea	*tropical japonica*	*badh2.1*	N/A
016	IRGC 38994	RA 7152	Bico Branco	Brazil	*Group V (aromatic)*	WT	<0.05
017	IRGC 26872	RA 7153	Binulawan	Philippines	*indica*	WT	<0.05
018	IRGC 45195	RA 7154	BJ 1	India	*aus*	WT	<0.05
019	IRGC 40275	RA 7155	BLACK GORA	India	*aus*	WT	<0.05
021	Clor 12248	RA 7157	BYAKKOKU Y 5006 SELN	Australia	*indica*	WT	<0.05
023	IRGC 50448	RA 7159	Canella De Ferro	Brazil	*tropical japonica*	WT	<0.05
027	IRGC 27869	RA 7163	Chahora 144	Pakistan	*tropical japonica*	WT	<0.05
028	IRGC 30238	RA 7164	Champa Tong 54	Thailand	*aus*	WT	<0.05
029	IRGC 56036	RA 7165	Chau	Vietnam	*indica*	WT	N/A
030	IRGC 10214	RA 7166	Chiem Chanh	Vietnam	*indica*	WT	<0.05
031	PI 431222	RA 7167	Chinese	China	*temperate japonica*	WT	<0.05
033	IRGC 17052	RA 7169	Chuan 4	Taiwan	*aus*	WT	<0.05
034	IRGC 3697	RA 7170	CO 25	India	*indica*	WT	N/A
035	IRGC 6331	RA 7171	CO18	India	*indica*	WT	<0.05
036	Clor 9675	RA 7172	Cs-M3	USA	*temperate japonica*	WT	N/A
037	IRGC 10658	RA 7173	Cuba 65	Cuba	*tropical japonica*	WT	<0.05
038	IRGC 5855	RA 7174	DA 5	Bangladesh	*tropical japonica*	WT	<0.05
041	IRGC 27630	RA 7177	Darmali	Nepal	*tropical japonica*	WT	<0.05
042	IRGC 8244	RA 7178	Davao	Philippines	*tropical japonica*	WT	<0.05
043	PI 279131	RA 7179	Dee Geo Woo Gen	Taiwan	*indica*	WT	<0.05

Supplemental Table 4.1 (continued)

044	IRGC 3686	RA 7180	Dhala Shaitta	Bangladesh	*aus*	WT	<0.05
045	PI 584607	RA 7181	DOM-SOFID	Iran	*Group V (aromatic)*	*badh2.1*	0.26
046	IRGC 3297	RA 7182	Dourado Agulha	Brazil	*tropical japonica*	WT	<0.05
049	IRGC 8839	RA 7185	DV85	Bangladesh	*aus*	WT	<0.05
050	IRGC 8555	RA 7186	DZ78	Bangladesh	*aus*	WT	<0.05
051	Clor 9738	RA 7187	Early Wataribune	Japan	*temperate japonica*	WT	<0.05
052	IRGC 8241	RA 7188	EH-1A-Chiu	Taiwan	*tropical japonica*	WT	N/A
053	IRGC 39261	RA 7189	Firooz	Iran	*Group V (aromatic)*	WT	<0.05
054	Clor 1344	RA 7190	Fortuna	USA	*tropical japonica*	WT	<0.05
055	IRGC 32300	RA 7191	Gerdeh	Iran	*Group V (aromatic)*	*badh2.1*	0.35
056	PI 597018	RA 7192	GEUMOBYEO	Korea	*temperate japonica*	WT	<0.05
057	IRGC 32303	RA 7193	Gharib	Iran	*indica*	WT	N/A
058	IRGC 58278	RA 7194	GHATI KAMMA NANGARHAR	Afghanistan	*aus*	WT	<0.05
059	IRGC 43394	RA 7195	Gogo Lempuk	Indonesia	*tropical japonica*	*badh2.1*	N/A
060	IRGC 43397	RA 7196	Gotak Gatik	Indonesia	*tropical japonica*	WT	<0.05
061	IRGC 51300	RA 7197	Guan-Yin-Tsan	China	*indica*	WT	<0.05
065	IRGC 1717	RA 7201	Honduras	Honduras	*tropical japonica*	WT	<0.05
066	IRGC 8240	RA 7202	Hsia-Chioh-Keh-Tu	Taiwan	*indica*	WT	<0.05
067	IRGC 8264	RA 7203	Hu Lo Tao	China	*temperate japonica*	WT	<0.05
069	IRGC 19642	RA 7205	IAC 25	Brazil	*tropical japonica*	WT	<0.05
071	IRGC 30416	RA 7207	IR 36	Philippines	*indica*	WT	<0.05
072	PI 312627	RA 7208	IR 8	Philippines	*indica*	WT	<0.05
073	GID 58752D	RA 7209	IRAT 177	French Guiana	*tropical japonica*	WT	<0.05
075	IRGC 17757	RA 7211	Jambu	Indonesia	*tropical japonica*	WT	<0.05
076	IRGC 11099	RA 7212	Jaya	India	*indica*	WT	<0.05
077	IRGC 9070	RA 7213	JC149	India	*indica*	WT	N/A
078	IRGC 6307	RA 7214	Jhona 349	India	*aus*	WT	<0.05
079	N/A	RA 7215	Jouiku 393G	Japan	*temperate japonica*	WT	<0.05
081	IRGC 45975	RA 7217	Kalamkati	India	*aus*	WT	<0.05
083	PI 403629	RA 7219	Kamenoo	Japan	*temperate japonica*	WT	<0.05
085	HO1195E	RA 7221	Kasalath	India	*aus*	WT	<0.05
087	IRGC 19972	RA 7223	Keriting Tingii	Indonesia	*tropical japonica*	*badh2.1*	0.47
088	IRGC 24224	RA 7224	Khao Gaew	Thailand	*aus*	WT	<0.05
089	IRGC 24225	RA 7225	Khao Hawm	Thailand	*tropical japonica*	WT	<0.05
090	IRGC 8180	RA 7226	Kiang-Chou-Chiu	Taiwan	*indica*	WT	<0.05
092	IRGC 3782	RA 7228	Kinastano	Philippines	*tropical japonica*	WT	<0.05
093	IRGC 12793	RA 7229	Kitrana 508	Madagascar	*Group V (aromatic)*	*badh2.1*	0.48
094	PI 330464	RA 7230	Koshihikari	Japan	*temperate japonica*	WT	<0.05
095	IRGC 2545	RA 7231	Kotobuki Mochi	Japan	*tropical japonica*	WT	<0.05
096	PI 597044	RA 7232	KU115	Thailand	*tropical japonica*	WT	N/A
097	IRGC 8195	RA 7233	Kun-Min-Tsieh-Hunan	China	*indica*	WT	<0.05
098	IRGC 66292	RA 7234	L-202	USA	*tropical japonica*	WT	<0.05
099	IRGC 14957	RA 7235	LAC 23	Liberia	*tropical japonica*	WT	<0.05
101	IRGC 66756	RA 7237	Lemont	USA	*tropical japonica*	WT	<0.05
102	IRGC 27762	RA 7238	Leung Pratew	Thailand	*temperate japonica*	WT	<0.05
103	IRGC 58286	RA 7239	Luk Takhar	Afghanistan	*temperate japonica*	WT	<0.05
105	IRGC12883	RA 7241	Mehr	Iran	*aus*	WT	<0.05
107	IRGC 25901	RA 7243	Miriti	Bangladesh	*tropical japonica*	WT	<0.05
108	IRGC 12048	RA 7244	Moroberekan	Guinea	*tropical japonica*	WT	<0.05
110	IRGC 6663	RA 7246	Mudgo	India	*indica*	WT	<0.05
111	IRGC 6264	RA 7247	N 22	India	*tropical japonica*	WT	<0.05
112	IRGC 6298	RA 7248	N12	India	*Group V (aromatic)*	*badh2.1*	0.47
113	IRGC 418	RA 7249	NORIN 20	Japan	*temperate japonica*	WT	<0.05
114	Clor 9459	RA 7250	Nova	USA	*temperate japonica*	WT	<0.05
115	IRGC 38696	RA 7251	NPE 835	Pakistan	*temperate japonica*	WT	<0.05
116	IRGC 38698	RA 7252	NPE 844	Pakistan	*tropical japonica*	WT	<0.05
117	IRGC 8179	RA 7253	O-LUEN-CHEUNG	Taiwan	*indica*	WT	<0.05
118	PI 439118	RA 7254	Oro	Chile	*temperate japonica*	WT	<0.05
119	PI 584668	RA 7255	Oryzica Llanos 5	Colombia	*indica*	WT	<0.05
120	PI 458474	RA 7256	OS6	Nigeria	*tropical japonica*	WT	<0.05
121	Clor 8	RA 7257	Ostiglia	Argentina	*temperate japonica*	WT	<0.05
122	IRGC 8261	RA 7258	Padi Kasalle	Indonesia	*tropical japonica*	WT	<0.05
125	IRGC 51400	RA 7261	Pao-Tou-Hung	China	*indica*	WT	<0.05
126	IRGC 8268	RA 7262	Pappaku	Taiwan	*indica*	WT	N/A
128	IRGC 5766	RA 7264	Pato De Gallinazo	Australia	*temperate japonica*	WT	<0.05
129	IRGC 8238	RA 7265	Peh-Kuh	Taiwan	*indica*	WT	<0.05
130	IRGC 8237	RA 7266	Peh-Kuh-Tsao-Tu	Taiwan	*indica*	WT	<0.05
131	IRGC 32399	RA 7267	Phudugey	Bhutan	*aus*	WT	<0.05
132	IRGC 8952	RA 7268	Rathuwee	Sri Lanka	*indica*	*badh2.1*	0.29
134	PI 433512	RA 7270	Romeo	Italy	*temperate japonica*	WT	<0.05
135	IRGC 15092	RA 7271	RT 1031-69	Zaire	*tropical japonica*	WT	<0.05
136	IRGC 8234	RA 7272	RTS12	Vietnam	*indica*	*badh2.1*	0.46
137	IRGC 8178	RA 7273	RTS14	Vietnam	*indica*	WT	<0.05
138	IRGC 8177	RA 7274	RTS4	Vietnam	*indica*	WT	<0.05
139	IRGC 1923	RA 7275	S4542A3-49B-2B12	USA	*tropical japonica*	WT	<0.05
140	Clor 9540	RA 7276	Saturn	USA	*temperate japonica*	WT	<0.05
141	IRGC 8260	RA 7277	Seratoes Hari	Indonesia	*indica*	WT	<0.05
143	Clor 1642	RA 7279	Shinriki	Japan	*temperate japonica*	WT	<0.05
146	IRGC 8242	RA 7282	Shuang-Chiang	Taiwan	*indica*	WT	N/A
147	IRGC 3967	RA 7283	Sinampaga Selection	Philippines	*tropical japonica*	WT	<0.05

Supplemental Table 4.1 (continued)

148	IRGC 5418	RA 7284	Sintane Diofor	Burkino Faso	*indica*	WT	<0.05
149	Clor 3337	RA 7285	SINAGUING	Philippines	*tropical japonica*	WT	<0.05
150	Clor 1658	RA 7286	Sultani	Egypt	*tropical japonica*	WT	<0.05
152	IRGC 6294	RA 7288	T 1	India	*aus*	WT	<0.05
154	IRGC 1107	RA 7290	TA HUNG KU	China	*temperate japonica*	WT	<0.05
155	IRGC 8194	RA 7291	Ta Mao Tsao	China	*temperate japonica*	WT	<0.05
156	PI 271672	RA 7292	Taichung Native 1	Taiwan	*indica*	WT	<0.05
157	PI 215936	RA 7293	Tainan Iku 487	Taiwan	*temperate japonica*	WT	<0.05
158	PI 366153	RA 7294	Taipei 309	Taiwan	*temperate japonica*	WT	<0.05
160	IRGC 32362	RA 7296	Tchampa	Iran	*Group V (aromatic)*	WT	<0.05
161	PI 536047	RA 7297	Teqing	China	*indica*	WT	<0.05
162	IRGC 237	RA 7298	TKM6	India	*indica*	WT	<0.05
163	PI 280681	RA 7299	Taducan	Philippines	*indica*	WT	<0.05
164	IRGC 3575	RA 7300	Tondok	Indonesia	*tropical japonica*	WT	<0.05
165	IRGC 43675	RA 7301	Trembese	Indonesia	*tropical japonica*	WT	<0.05
167	Clor 9881	RA 7303	B6616A4-22-Bk-5-4	USA	*tropical japonica*	WT	<0.05
169	Clor 5309	RA 7305	WC 6	China	*tropical japonica*	*badh2.1*	0.53
170	PI 612439	RA 7306	Wells	USA	*tropical japonica*	WT	<0.05
171	PI 629016	RA 7307	Zhe 733	China	*indica*	WT	<0.05
172	PI 602624	RA 7308	Zhenshan 2	China	*indica*	WT	<0.05
173	IRGC 117274	RA 7309	Nipponbare	Japan	*temperate japonica*	WT	<0.05
174	IRIS 2254730	RA 7310	Azucena	Philippines	*tropical japonica*	*badh2.1*	0.31
391	Clor 9483	RA 7527	Della	USA	*tropical japonica*	*badh2.1*	1.43
	IRGC 52266	RA 2221	Ambemohar 1	India	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 78469	RA 2225	Basmati Mehtrah	India	*Group V (aromatic)*	WT	<0.05
	IRGC 79560	RA 2226	Basmati Bahar (shorter grain)	India	*Group V (aromatic)*	*badh2.1*	0.27
	IRGC 27815	RA 2232	Basmati 213	Pakistan	*indica*	WT	N/A
	IRGC 27822	RA 2237	Basmati 370B	Pakistan	*Group V (aromatic)*	WT	N/A
		RA 2239	Basmati 150	India	*Group V (aromatic)*	WT	N/A
	IRGC 53827	RA 2248	BPT1235	India	*Group V (aromatic)*	*badh2.1*	N/A
		RA 2250	Bogi Joha	India	*indica*	WT	N/A
		RA 2253	Barhail	India	*Group V (aromatic)*	*badh2.1*	N/A
		RA 2254	Blomberg	India	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 45298	RA 2258	Chinisakkar	India	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 43805	RA 2259	Chinigura	Bangladesh	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 78476	RA 2262	Dulhamiya	India	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 8628	RA 2264	DD66	Bangladesh	*Group V (aromatic)*	*badh2.1*	N/A
CCSHAU		RA 2271	Haryana Basmati1	India	*Group V (aromatic)*	*badh2.1*	N/A
CCSHAU		RA 2277	HKR93-401	India	*Group V (aromatic)*	*badh2.1*	N/A
CCSHAU		RA 2278	Tararori Basmati	India	*Group V (aromatic)*	*badh2.1*	0.68
	IRGC 78422	RA 2279	Pusa basmati-1	India	*Group V (aromatic)*	*badh2.1*	N/A
CCSHAU		RA 2280	HBC-19	India	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 27748	RA 4878	Khao Dawk Mali 105	Thailand	*indica*	*badh2.1*	1.21
	IRIS 2254721	RA 4891	FR13 A	India	*aus*	WT	<0.05
	IRGC 27513	RA 4984	Dholi Boro	Bangladesh	*indica*	WT	<0.05
		RA 5133	IR64	Philippines	*indica*	WT	<0.05
JP13	collected in Indo	RA 5756	P. Pulut Longbanga	Indonesia	*tropical japonica*	*badh2.1*	N/A
JP152	collected in Indo	RA 5895	P. Telengusan	Indonesia	*tropical japonica*	*badh2.1*	N/A
JP176	collected in Indo	RA 5919	P. Ciu	Indonesia	*tropical japonica*	*badh2.1*	N/A
JP178	collected in Indo	RA 5921	P. Ikeng	Indonesia	*tropical japonica*	*badh2.1*	N/A
JP186	collected in Indo	RA 5929	P. Pampang	Indonesia	*indica*	*badh2.1*	N/A
JP187	collected in Indo	RA 5930	P. Gadis	Indonesia	*tropical japonica*	*badh2.1*	N/A
JP55	collected in Indo	RA 5798	P. Pulut Mbau	Indonesia	*tropical japonica*	*badh2.1*	N/A
		RA 8126	Q74	Malaysia	*indica*	*badh2.1*	0.70
		RA 8127	Nang Thom Cho Dao	Vietnam	*indica*	*badh2.1*	0.28
		RA 8130	RD33	Thailand	*indica*	*badh2.1*	0.66
		RA 8131	PTT1	Thailand	*indica*	*badh2.1*	0.21
		RA 8132	RD6	Thailand	*indica*	*badh2.1*	0.43
		RA 8135	OM3536	Vietnam	*indica*	*badh2.1*	0.34
		RA 8136	Basmati 2000	Pakistan	*Group V (aromatic)*	*badh2.1*	0.26
		RA 8141	Nho Thom	Vietnam	*indica*	*badh2.1*	0.63
		RA 8147	Chao do	Laos	*indica*	*badh2.1*	0.19
		RA 8155	HNN	Laos	*indica*	*badh2.1*	0.36
	IRGC 6148	RA 8157	AMBEMOHOR 157	India	*Group V (aromatic)*	*badh2.1*	0.84
	IRGC 20310	RA 8158	ARC 6011	India	*Group V (aromatic)*	*badh2.1*	1.13
	IRGC 27781	RA 8159	BARA 413	Pakistan	*Group V (aromatic)*	*badh2.1*	0.89
	IRGC 29258	RA 8160	BASFUL 714	Bangladesh	*Group V (aromatic)*	*badh2.1*	0.90
	IRGC 10330	RA 8162	BASMATI 6311	Pakistan	*Group V (aromatic)*	*badh2.1*	1.41
	IRGC 9027	RA 8163	BASMATI C 622	Pakistan	*Group V (aromatic)*	*badh2.1*	0.82
	IRGC 5857	RA 8169	DA13	Bangladesh	*Group V (aromatic)*	*badh2.1*	1.70
	IRGC 32291	RA 8171	DOMSIAH	Iran	*Group V (aromatic)*	*badh2.1*	1.68
	IRGC 27953	RA 8174	HARANDI 379	Pakistan	*Group V (aromatic)*	*badh2.1*	1.21
	IRGC 32329	RA 8175	SADRI	Iran	*Group V (aromatic)*	*badh2.1*	1.08
	IRGC 42736	RA 8176	ARC 13523	India	*Group V (aromatic)*	WT	<0.05
	IRGC 20330	RA 8177	ARC 6070	India	*Group V (aromatic)*	WT	<0.05
	IRGC 27856	RA 8178	BEGUMI 302	Pakistan	*Group V (aromatic)*	WT	<0.05
	IRGC 32390	RA 8181	FARANGEY	Bhutan	*Group V (aromatic)*	WT	<0.05
	IRGC 58288	RA 8182	MAHIN KUNDUZ	Afghanistan	*Group V (aromatic)*	WT	<0.05
	IRGC 4802	RA 8183	RANDHANIPAGAL(SCENTED)	China	*temperate japonica*	WT	<0.05
	IRGC 9026	RA 8185	BASMATI 370	India	*Group V (aromatic)*	*badh2.1*	0.87

Supplemental Table 4.1 (continued)

	IRGC 32985	RA 8186	BOKEHMWE	Myanmar	*indica*	*badh2.1*	1.60
	IRGC 16107	RA 8187	GUOR LAO	Vietnam	*indica*	*badh2.1*	1.32
	IRGC 9092	RA 8188	JC220	India	*Group V (aromatic)*	*badh2.1*	0.47
	IRGC 33357	RA 8189	MA WAINE OHN	Myanmar	*indica*	*badh2.1*	1.00
	IRGC 33571	RA 8191	PAWSANHMWE	Myanmar	*indica*	*badh2.1*	1.41
	IRGC 78747	RA 8193	TAINUNG SEN 20	Taiwan	*indica*	*badh2.1*	0.68
	IRGC 233103	RA 8195	Aus Basmati	India	*Group V (aromatic)*	*badh2.1*	0.43
	IRGC 385453	RA 8200	Basmati 410	India	*Group V (aromatic)*	*badh2.1*	0.63
	IRGC 385785	RA 8201	Basmati Lal	Pakistan	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 412772	RA 8206	Basmati 802	Pakistan	*Group V (aromatic)*	*badh2.1*	1.21
	IRGC 412852	RA 8208	Karnal Basmati	Pakistan	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 584591	RA 8222	YUN AROMATIC GLUTINOUS	China	*indica*	*badh2.1*	N/A
	IRGC 595927	RA 8224	JASMINE 85	USA	*indica*	*badh2.1*	0.38
	GSOR 310319	RA 8230	BC5-55	India	*indica*	*badh2.1*	N/A
	GSOR 311134	RA 8231	Sathi Basmati	Pakistan	*Group V (aromatic)*	*badh2.1*	0.44
	GSOR 311216	RA 8232	Basmati 802	Pakistan	*Group V (aromatic)*	*badh2.1*	1.21
	IRGC 77952	RA 8362	Bidor	Malaysia	*tropical japonica*	*badh2.1*	N/A
	IRGC 77954	RA 8364	Coreng	Malaysia	*tropical japonica*	*badh2.1*	N/A
	IRGC 115287	RA 8395	Ble Ma Mua	Vietnam	*tropical japonica*	WT	N/A
	IRGC 9877	RA 8503	Kalijira	Bangladesh	*Group V (aromatic)*	*badh2.1*	0.20
	IRGC 66247	RA 8514	Sadri Ghermes	Iran	*Group V (aromatic)*	*badh2.1*	N/A
	IRGC 78480	RA 8516	Jasmine Scented	Thailand	*indica*	*badh2.1*	0.84
	IRGC 87902	RA 8521	Somaly Krar-oob	Cambodia	*indica*	*badh2.1*	N/A
	IRGC 27798	F14	Basmati 1_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.73
	IRGC 27815	F16	Basmati 213_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.26
	IRGC 27819	F17	Basmati 334_IRRI	Pakistan	*Group V (aromatic)*	WT	<0.05
	IRGC 27822	F18	Basmati 370B_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.73
	IRGC 27824	F19	Basmati 372A_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.67
	IRGC 27869	F20	Chahora 144_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.39
	IRGC 30238	F21	Champa Tong 54_IRRI	Thailand	*indica*	*badh2.1*	0.27
	IRGC 9091	F3	JC1_IRRI	India	*Group V (aromatic)*	*badh2.1*	0.91
	IRGC 78470	F33	Basmati sathi_IRRI	India	*Group V (aromatic)*	*badh2.1*	0.35
	IRGC 5999	F38	Pankhari 203_IRRI	India	*Group V (aromatic)*	*badh2.1*	0.46
	IRGC 7937	F4	Pratao_IRRI	Brazil	*indica*	WT	<0.05
	IRGC 42469	F58	ARC 13829_IRRI	India	*Group V (aromatic)*	WT	<0.05
	IRGC 9036	F61	Basmati 6113_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.40
	IRGC 9030	F63	Basmati 5836_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.73
	IRGC 27805	F75	Basmati 122_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*	0.16
	IRGC 58930	F79	Chhote Dhan_IRRI	Nepal	*Group V (aromatic)*	WT	<0.05
	IRGC 78469	F80	Basmati mehtrah_IRRI	India	*Group V (aromatic)*	WT	<0.05
	IRGC 6245	AUS1	DA16	Bangladesh	*aus*	WT	N/A
	IRGC 6307	AUS2	JHONA 349	India	*aus*	WT	<0.05
	IRGC 6422	AUS3	DA8	Bangladesh	*aus*	WT	N/A
	IRGC 7702	AUS4	KALUBALA VEE	Sri Lanka	*aus*	WT	N/A
	IRGC 9069	AUS5	JC148	India	*aus*	WT	N/A
	IRGC 25835	AUS6	BAILAM	Bangladesh	*aus*	WT	N/A
	IRGC 25852	AUS7	DUMAI	Bangladesh	*aus*	WT	N/A
	IRGC 32300	AUS8	GERDEH	Iran	*aus*	WT	N/A
	IRGC 34712	AUS9	LAKHI PURI	Bangladesh	*aus*	WT	N/A

Fragrant Accessions Lacking Known Derived *BADH2* Alleles:

Code	Gene Bank #	Cornell RA #	Accession Name	Country of Origin	Subpopulation	BADH2 allele	[2AP]
ANF4	IRGC 28473	RA 8203	HSIANG-KENG-NUO	China	*tropical japonica*	*badh2.2*	0.66
ANF10	IRGC 69874	RA 8515	RASOMOTRAFOTSY	Madagascar	*tropical japonica*	*badh2.3*	0.74
ANF3	IRGC 27348	RA 8507	PARE BAINE PULUT	Indonesia	*tropical japonica*	*badh2.4*	0.59
ANF9	IRGC 52787	RA 8513	VISHUNPARAG	India	*aus*	*badh2.5*	0.41
ANF1	IRGC 13453	RA 8505	PADI WANGI	Malaysia	*indica*	*badh2.6*	0.70
ANF6	IRGC 34975	RA 8510	KATAK TARA	India	*aus*	*badh2.7*	0.37
ANF7	IRGC 36394	RA 8511	SUWANDA AL	Sri Lanka	*aus*	*badh2.7*	0.30
ANF8	IRGC 44204	RA 8512	UPRB28	India	*aus*	*badh2.7*	0.42
ANF11	IRGC 82408	RA 8518	BASMATI BAHAR(L-GRAIN)	India	*aus*	*badh2.7*	0.38
ANF12	IRGC 83308	RA 8519	BASMATI	Nepal	*aus*	*badh2.7*	0.32
ANF13	IRGC 83309	RA 8520	BASMATI	Nepal	*aus*	*badh2.7*	0.77
ANF5	IRGC 33858	RA 8509	YANGON SABA	Myanmar	*Group V (aromatic)*	*badh2.8*	0.31
F22	IRGC 32960	RA 8508	Balugyun_IRRI	Myanmar	*Group V (aromatic)*	*badh2.8*	0.42
SR5	IRGC 97793	RA 8524	Paw San Hmwe_IRRI	Myanmar	*Group V (aromatic)*	*badh2.8*	0.43
			Paw San Hmwe_METAPHOR	Myanmar		*badh2.8*	0.49
	IRGC 33064	RA 8172	EMATA LONGYU	Myanmar	*Group V (aromatic)*	*badh2.8*	0.14
ANF2	IRGC 18438	RA 8506	PANDAN WANGI	Indonesia	*tropical japonica*	*badh2.9*	0.29
PW1	N/A	DNA only	Pandan Wangi	Indonesia	*tropical japonica*	*badh2.9*	0.19
SR8	N/A	DNA only	Pandan Wangi	Indonesia	*tropical japonica*	*badh2.9*	0.14
SR9	N/A	DNA only	Pandan Wangi	Indonesia	*tropical japonica*	*badh2.9*	0.18
	IRGC 79569	RA 8517	Mentik Wangi	Indonesia	*tropical japonica*	*badh2.9*	N/A
		RA 8151	Pandan Wangi8	Indonesia	*tropical japonica*	*badh2.9*	0.08
	IRGC 78746	RA 8192	TAINUNG 72	Taiwan	*tropical japonica*	*badh2.10*	0.35
	IRGC 439612	RA 8215	Shiratama	Japan	*tropical japonica*	*badh2.10*	N/A
ANF14	IRGC 90687	RA 8522	KHAU TAN LUONG	Vietnam	*tropical japonica*	WT	0.79
	LG6732	RA 8156/K2	Kai Noi Leuang	Laos	*tropical japonica*	WT	0.76

Supplemental Table 4.2: Gene Haplotype for All 242 *O. sativa* Accessions

	Accession Name	Subpopulation	IRGC #	ID#	Gene Haplotype #	Gene Haplotype Cluster	Polymorphisms' Pseudomolecule Position on Chromosome 8 (TIGR v.5) 20246946 20246960 20246997 20247005 20247019 20247029 20247035 20247492 20247498 20247555 20247558 20247695 20247705 20247730 20247816 20247965 20248041 20248208 20248239 20248251 20248295 20248469 20248945 20248977 20249064 20249216 20249245 20249288 20249270 20249280** 20249314 20249318 20249333 20250015 20250017 20250019 20250027 20250106 20250176 20250311 20250346 20250466 20250555 20250563 20250605 20250913 20251108 20251114 20251137 20251469 20251577 20252057 20252365 20252967
1	ARC 6070	Group V	IRGC 20330	8177	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
2	ARC 13829	Group V	IRGC 42469	F56	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
3	Basmati 370D	Group V	IRGC 27022	2237	1	Jap_GI	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
4	Basmati 150	Group V		2239	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
5	Chhote Dhan	Group V	IRGC 58930	F79	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
6	Champa	Group V	IRGC 32362	160	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
7	ARC 13523	Group V	IRGC 42736	8176	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
8	Farangey	Group V	IRGC 32390	8181	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
9	NORIN 20	temperate japonica	IRGC 418	113	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
10	Hu Lu Tau	temperate japonica	IRGC 8264	67	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
11	GEUMOBYEO	temperate japonica	PI 597118	56	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
12	Kameroo	temperate japonica	PI 403529	83	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
13	Jouiku 393G	temperate japonica		79	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
14	Kochihikari	temperate japonica	PI 330464	94	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
15	Ta Mae Tsao	temperate japonica	IRGC 8194	155	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
16	Ostiglia	temperate japonica	Clor 8	121	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
17	NPE 835	temperate japonica	IRGC 38696	115	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
18	Remer	temperate japonica	PI 433512	134	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
19	Cs-M3	temperate japonica	Clor 9675	36	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
20	Pato De Galinazo	temperate japonica	IRGC 5766	128	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
21	TA HUNG KU	temperate japonica	IRGC 1107	154	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
22	Laung Pratew	temperate japonica	IRGC 27762	102	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
23	Agostano	temperate japonica	IRGC 3135	1	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
24	Cro	temperate japonica	PI 439118	118	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
25	Gotak Gatik	tropical japonica	IRGC 43397	60	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
26	Trembese	tropical japonica	IRGC 43675	165	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
27	Darmai	tropical japonica	IRGC 27630	41	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
28	KU115	tropical japonica	PI 597044	96	1	Jap_GI	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
29	L-202	tropical japonica	IRGC 66292	98	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
30	DA 5	tropical japonica	IRGC 5855	38	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
31	Kotobuki Mochi	tropical japonica	IRGC 2545	95	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
32	Moroberekan	tropical japonica	IRGC 12048	108	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
33	Canella De Ferro	tropical japonica	IRGC 50448	23	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
34	Chahora 144	tropical japonica	IRGC 27869	27	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 0 C G A A A G A C A G C G T T C T T A C C 4 G C
35	Mirili	tropical japonica	IRGC 25901	107	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
36	NPF 814	tropical japonica	IRGC 38698	116	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
37	Kinastano	tropical japonica	IRGC 3782	92	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
38	IRAT 177	tropical japonica	GID 56752D	73	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 0 C G A A A G A C A G C G T T C T T A C C 4 G C
39	Honduras	tropical japonica	IRGC 1717	65	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
40	CS6	tropical japonica	PI 458474	120	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A A A G A C A G C G T T C T T A C C 4 G C
41	Douraco Agulha	tropical japonica	IRGC 3297	46	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
42	IAC 25	tropical japonica	IRGC 19642	69	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
43	LAC 23	tropical japonica	IRGC 14957	99	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
44	RT 1031-69	tropical japonica	IRGC 15092	135	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
45	Cuba 65	tropical japonica	IRGC 10658	37	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
46	Davao	tropical japonica	IRGC 8244	42	1	Jap_CH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C C A A A G A C A G C G T T C T T A C C 4 G C
47	Elle Ma Mua	tropical japonica	IRGC 115287	8395	1	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A A A G A C A G C G T T C T T A C C 4 G C
48	JC1	Group V	IRGC 9091	F3	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
49	BASFUL 714	Group V	IRGC 29258	8160	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
50	HNN	indica		8165	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
51	Chao co	indica		8147	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
52	WC 6	tropical japonica	Clor 5309	169	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
53	Keriting Tingii	tropical japonica	IRGC 19972	97	2	Jap_GI	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 2 C G A T T T T A C A G C G T T C T T A C C 4 G C
54	Blomberg	Group V		2254	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C
55	Chini sakkar	Group V	IRGC 45298	2256	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C
56	TARAORI	Group V		2278	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A I I T I A C A G C G T I C T T A C C 4 G C
57	Pusa basmati 1	Group V	IRGC 78422	2279	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C
58	HBC-19	Group V		2280	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C
59	HARANDI 379	Group V	IRGC 27953	8174	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C
60	Basmati 802	Group V	GSOR 311216	8232	2	Jap_GH	G T C G C A A G C C T G T 2 C T A G 2 C T T T C C T T 4 C G A T T T T A C A G C G T T C T T A C C 4 G C

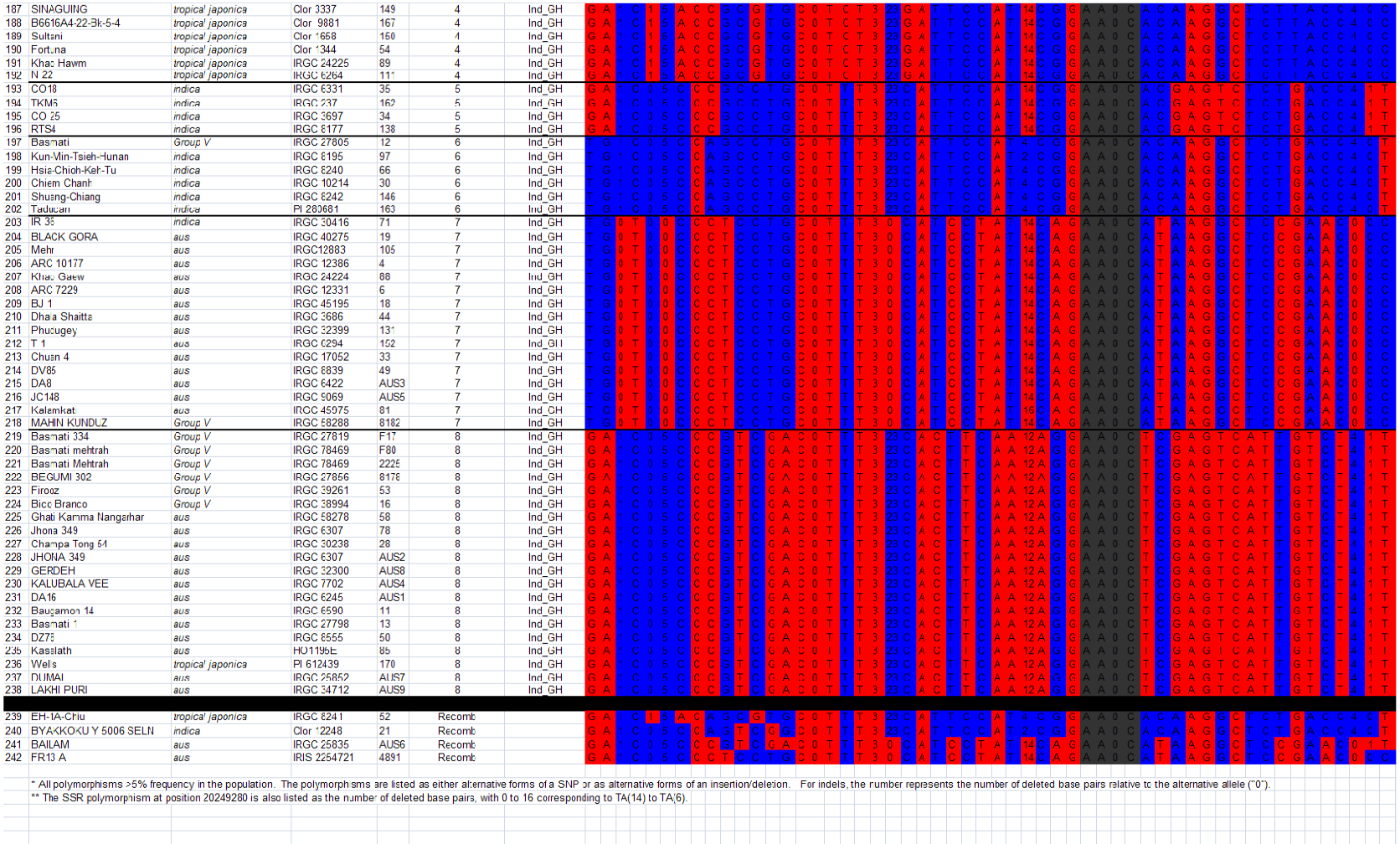
Supplemental Table 4.2 (continued)

61	HKR93(401)	Group V		2277	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
62	PTT1	indica		8131	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
63	Yun Aromatic Glutinous	indica	IRGC 584591	8222	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
64	Somaly Krar-oob	indica	IRGC 87902	8521	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
65	Jasmine Scented	indica	IRGC 78480	8516	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
66	Nang Thom Cho Dao	indica		8127	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
67	MA WAINE OHN	indica	IRGC 33357	8189	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
68	PAWSANHMWE	indica	IRGC 33571	8191	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
69	RTS12	indica	IRGC 8234	136	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
70	GUOR LAO	indica	IRGC 16107	8187	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
71	Q74	indica		8126	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
72	P. Ciu	indica		J186	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
73	Champa long 54	indica	IRGC 30238	F21	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
74	Rathuwee	indica	IRGC 8952	132	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
75	BOKEHMWE	indica	IRGC 32985	8186	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
76	Khao Dawk Mali 105	indica	IRGC 27748	4878	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
77	JASMINE 85	indica	IRGC 595927	8224	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
78	TAINUNG SEN 20	indica	IRGC 78747	8193	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
79	Nho Thom	indica		8141	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
80	RD33	indica		8130	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
81	RD6	indica		8132	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
82	OM3536	indica		8135	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
83	P. Gacis	tropical japonica		J55	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
84	P. Fulut Mbau	tropical japonica		J176	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
85	Ambemohar 1	Group V	IRGC 52266	2221	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
86	JC220	Group V	IRGC 9092	8188	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
87	AMBEMOHOR 157	Group V	IRGC 6148	8157	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
88	Sadri Ghermes	Group V	IRGC 66247	8514	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
89	Gerdeh	Group V	IRGC 32300	55	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
90	Basmati bahar	Group V	IRGC 79560	2226	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
91	ARC 6011	Group V	IRGC 20310	8158	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
92	Basmati 6113	Group V	IRGC 9036	F61	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
93	SADRI	Group V	IRGC 32329	8175	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
94	ARC 10352	Group V	IRGC 12440	5	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
95	Dulhemiya	Group V	IRGC 78476	2262	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
96	DOMSIAH	Group V	IRGC 32291	8171	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
97	DA13	Group V	IRGC 6857	8169	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
98	Kalijira	Group V	IRGC 9877	8503	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
99	Pankhari 203	Group V	IRGC 5999	F38	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
100	DOM-SOFID	Group V	PI 584607	45	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
101	Basmati 1	Group V	IRGC 27798	F14	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
102	Basmati 217	Group V	IRGC 53637	14	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
103	Basmati 213	Group V	IRGC 27815	F16	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
104	Basmati 370B	Group V	IRGC 27822	F18	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
105	Basmati sathi	Group V	IRGC 78470	F33	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
106	Kitrana 508	Group V	IRGC 12793	93	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
107	N12	Group V	IRGC 6298	112	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
108	BPT1235	Group V	IRGC 53827	2248	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
109	Barhai 1	Group V		2253	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
110	Chiri guri	Group V	IRGC 43805	2259	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
111	D66	Group V	IRGC 8628	2264	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
112	BASMATI 6311	Group V	IRGC 10330	8162	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
113	DASMATI C 622	Group V	IRGC 9027	0160	2	Jap_GII	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
114	Aus Basmati	Group V	IRGC 233103	8195	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
115	Chanora 144	Group V	IRGC 27869	F20	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
116	Basmati 5836	Group V	IRGC 9030	F63	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
117	BASMATI 370	Group V	IRGC 9026	8185	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
118	Basmati Lal	Group V	IRGC 385785	8201	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
119	BARA 413	Group V	IRGC 27781	8159	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
120	Sathi Basmati	Group V	GSOR 311134	8231	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
121	Basmati 2000	Group V		0136	2	Jap_GII	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
122	Basmati 372A	Group V	IRGC 27824	F19	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC
123	HKR 228	Group V		2271	2	Jap_GH	TGTCGCAAGCCTGTCCTAGCTTTCCTTACGATTGTACAGCGTTCTTACCACC

Supplemental Table 4.2 (continued)

124	Basmati 122	Group V	IRGC 27805	F75	2	Jap_GH	
125	Basmati 410	Group V	IRGC 385453	8200	2	Jap_GH	
126	Basmati 802	Group V	IRGC 412772	8206	2	Jap_GH	
127	Kamal Basmat	Group V	IRGC 412852	8208	2	Jap_GH	
128	BC5-55	indica	GSOR 310319	8230	2	Jap_GH	
129	P. Atok	tropical japonica		3152	2	Jap_GH	
130	P. Ikeng	tropical japonica		3187	2	Jap_GH	
131	P. Telengusan	tropical japonica		3178	2	Jap_GH	
132	Aras	tropical japonica	IRGC 43325	7	2	Jap_GH	
133	Beonjo	tropical japonica	IRGC 55457	15	2	Jap_GH	
134	P. Pulut Longbanga	tropical japonica		3113	2	Jap_GH	
135	Azucena	tropical japonica	IRIS 2254730	7737	2	Jap_GH	
136	Bidor	tropical japonica	IRGC 77952	8362	2	Jap_GH	
137	Ccreng	tropical japonica	IRGC 77954	8364	2	Jap_GH	
138	Asse Y Pung	tropical japonica	IRGC 6949	8	2	Jap_GH	
139	Della	tropical japonica	Clor 9483	391	2	Jap_GH	
140	Gogo Lempuk	tropical japonica	IRGC 43394	59	2	Jap_GH	
141	Chirese	temperate japonica	PI 431222	31	3	Jap_GH	
142	Early Wataribune	temperate japonica	Clor 9738	51	3	Jap_GH	
143	Luk Takhar	temperate japonica	IRGC 58286	103	3	Jap_GH	
144	Shinriki	temperate japonica	Clor 1642	143	3	Jap_GH	
145	Tainan Iku 487	temperate japonica	PI 215936	157	3	Jap_GH	
146	Taipei 309	temperate japonica	PI 366153	158	3	Jap_GH	
147	Nipponbare	temperate japonica	IRGC 117274	173	3	Jap_GH	
148	Randhaniosgal (scerted)	temperate japonica	IRGC 4802	8183	3	Jap_GH	
149	Daber	tropical japonica	IRGC 33904	9	3	Jap_GH	
150	Gharib	indica	IRGC 32303	57	4	Ind_GH	
151	Papoaku	indica	IRGC 8268	126	4	Ind_GH	
152	JC149	indica	IRGC 9070	77	4	Ind_GH	
153	Chau	indica	IRGC 56036	29	4	Ind_GH	
154	Zhe 733	indica	PI 529016	171	4	Ind_GH	
155	Guan-Yin-Tsan	indica	IRGC 51300	61	4	Ind_GH	
156	Kiarg-Chou-Chiu	indica	IRGC 8180	90	4	Ind_GH	
157	O-LUEN-CHEUNG	indica	IRGC 8179	117	4	Ind_GH	
158	Dee Geo Woo Gen	indica	PI 279131	43	4	Ind_GH	
159	Peh-Kuh-Tsao-Tu	indica	IRGC 8237	130	4	Ind_GH	
160	Taichung Native 1	indica	PI 271672	156	4	Ind_GH	
161	Teqing	indica	PI 536047	161	4	Ind_GH	
162	Jaya	indica	IRGC 11099	76	4	Ind_GH	
163	Basmati 213	indica	IRGC 27815	2232	4	Ind_GH	
164	Binulawan	indica	IRGC 26872	17	4	Ind_GH	
165	IR 8	indica	PI 312627	72	4	Ind_GH	
166	IR64	indica		5133	4	Ind_GH	
167	RTS14	indica	IRGC 8178	137	4	Ind_GH	
168	Oryzica Llanos 5	indica	PI 584668	119	4	Ind_GH	
169	Seratoee Hari	indica	IRGC 8260	141	4	Ind_GH	
170	Sintane Diofor	indica	IRGC 5418	148	4	Ind_GH	
171	Zhenshan 2	indica	PI 502624	172	4	Ind_GH	
172	Pao-Tou-Hung	indica	IRGC 51400	125	4	Ind_GH	
173	Ai-Chiao-Hong	indica	IRGC 51250	3	4	Ind_GH	
174	Peh-Kuh	indica	IRGC 8238	129	4	Ind_GH	
175	Dhoi Boro	indica	IRGC 27513	4984	4	Ind_GH	
176	Pratao	indica	IRGC 7937	F4	4	Ind_GH	
177	Beg Joha	indica		2250	4	Ind_GH	
178	Mudgo	indica	IRGC 6663	110	4	Ind_GH	
179	Nova	temperate japonica	Clor 9459	114	4	Ind_GH	
180	Saturn	temperate japonica	Clor 9540	140	4	Ind_GH	
181	Jambu	tropical japonica	IRGC 17757	75	4	Ind_GH	
182	Tondok	tropical japonica	IRGC 3575	164	4	Ind_GH	
183	Ped. Kasalle	tropical japonica	IRGC 8261	122	4	Ind_GH	
184	Lemont	tropical japonica	IRGC 66756	101	4	Ind_GH	
185	S4542A3-49B-2B12	tropical japonica	IRGC 1923	139	4	Ind_GH	
186	Sinampaga Selection	tropical japonica	IRGC 3967	147	4	Ind_GH	

Supplemental Table 4.2 (continued)



187	SINAGUING	*tropical japonica*	Clor 3337	149	4	Ind_GH	
188	B6616A4-22-3k-5-4	*tropical japonica*	Clor 9881	167	4	Ind_GH	
189	Sultani	*tropical japonica*	Clor 1658	150	4	Ind_GH	
190	Fortuna	*tropical japonica*	Clor 1344	54	4	Ind_GH	
191	Khao Hawm	*tropical japonica*	IRGC 24225	89	4	Ind_GH	
192	N 22	*tropical japonica*	IRGC 6264	111	4	Ind_GH	
193	CO18	*indica*	IRGC 6331	35	5	Ind_GH	
194	TKM6	*indica*	IRGC 237	162	5	Ind_GH	
195	CO 25	*indica*	IRGC 3697	34	5	Ind_GH	
196	RTS4	*indica*	IRGC 8177	138	5	Ind_GH	
197	Basmati	*Group V*	IRGC 27805	12	6	Ind_GH	
198	Kun-Min-Tsieh-Hunan	*indica*	IRGC 8195	97	6	Ind_GH	
199	Hsia-Chioh-Keh-Tu	*indica*	IRGC 8240	66	6	Ind_GH	
200	Chiem Chanh	*indica*	IRGC 10214	30	6	Ind_GH	
201	Shuang-Chiang	*indica*	IRGC 8242	146	6	Ind_GH	
202	Taducan	*indica*	PI 280681	163	6	Ind_GH	
203	IR 36	*indica*	IRGC 30416	71	7	Ind_GH	
204	BLACK GORA	*aus*	IRGC 40275	19	7	Ind_GH	
205	Mehr	*aus*	IRGC12883	105	7	Ind_GH	
206	ARC 10177	*aus*	IRGC 12386	4	7	Ind_GH	
207	Khao Gaew	*aus*	IRGC 24224	88	7	Ind_GH	
208	ARC 7229	*aus*	IRGC 12331	6	7	Ind_GH	
209	BJ 1	*aus*	IRGC 45195	18	7	Ind_GH	
210	Dhaia Shaitta	*aus*	IRGC 3686	44	7	Ind_GH	
211	Phucugey	*aus*	IRGC 32399	131	7	Ind_GH	
212	T 1	*aus*	IRGC 6294	152	7	Ind_GH	
213	Chuan 4	*aus*	IRGC 17052	33	7	Ind_GH	
214	DV85	*aus*	IRGC 8839	49	7	Ind_GH	
215	DA8	*aus*	IRGC 6422	AUS3	7	Ind_GH	
216	JC148	*aus*	IRGC 9069	AUS5	7	Ind_GH	
217	Kalamkati	*aus*	IRGC 45975	81	7	Ind_GH	
218	MAHIN KUNDUZ	*Group V*	IRGC 58288	8182	7	Ind_GH	
219	Basmati 334	*Group V*	IRGC 27819	F17	8	Ind_GH	
220	Basmati mehtrah	*Group V*	IRGC 78469	F80	8	Ind_GH	
221	Basmati Mehtrah	*Group V*	IRGC 78469	2225	8	Ind_GH	
222	BEGUMI 302	*Group V*	IRGC 27856	8178	8	Ind_GH	
223	Firooz	*Group V*	IRGC 39261	53	8	Ind_GH	
224	Bico Branco	*Group V*	IRGC 38994	16	8	Ind_GH	
225	Ghati Kamma Nangarhar	*aus*	IRGC 58278	58	8	Ind_GH	
226	Jhona 349	*aus*	IRGC 6307	78	8	Ind_GH	
227	Champa Tong 64	*aus*	IRGC 30238	28	8	Ind_GH	
228	JHONA 349	*aus*	IRGC 6307	AUS2	8	Ind_GH	
229	GERDEH	*aus*	IRGC 32300	AUS8	8	Ind_GH	
230	KALUBALA VEE	*aus*	IRGC 7702	AUS4	8	Ind_GH	
231	DA16	*aus*	IRGC 6245	AUS1	8	Ind_GH	
232	Baugamon 14	*aus*	IRGC 6590	11	8	Ind_GH	
233	Basmati 1	*aus*	IRGC 27798	13	8	Ind_GH	
234	DZ78	*aus*	IRGC 6555	50	8	Ind_GH	
235	Kasalath	*aus*	HO1195E	85	8	Ind_GH	
236	Wells	*tropical japonica*	PI 612439	170	8	Ind_GH	
237	DJIMAI	*aus*	IRGC 25852	AUS7	8	Ind_GH	
238	LAKHI PURI	*aus*	IRGC 34712	AUS9	8	Ind_GH	
239	EH-1A-Chiu	*tropical japonica*	IRGC 6241	52	Recomb		
240	BYAKKOKU Y 5006 SELN	*indica*	Clor 12248	21	Recomb		
241	BAILAM	*aus*	IRGC 25835	AUS6	Recomb		
242	FR13 A	*aus*	IRIS 2254721	4891	Recomb		

* All polymorphisms >5% frequency in the population. The polymorphisms are listed as either alternative forms of a SNP or as alternative forms of an insertion/deletion. For indels, the number represents the number of deleted base pairs relative to the alternative allele ("0").

** The SSR polymorphism at position 20249280 is also listed as the number of deleted base pairs, with 0 to 16 corresponding to TA(14) to TA(6).

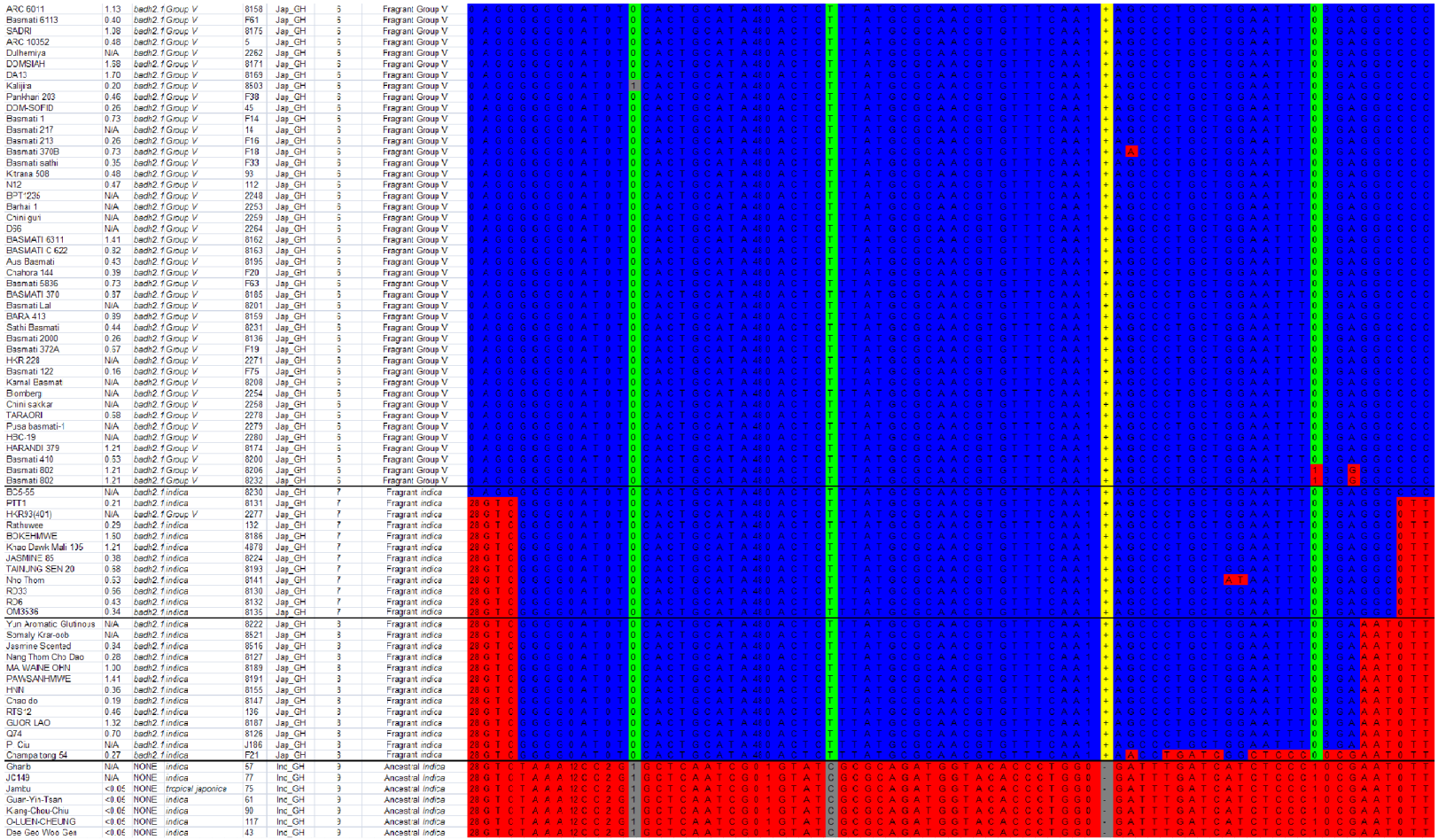
Supplemental Table 4.3: Extended Haplotypes for All 242 *O. sativa* Accessions and the Heterozygous Wild Accession

Supplemental Table S3

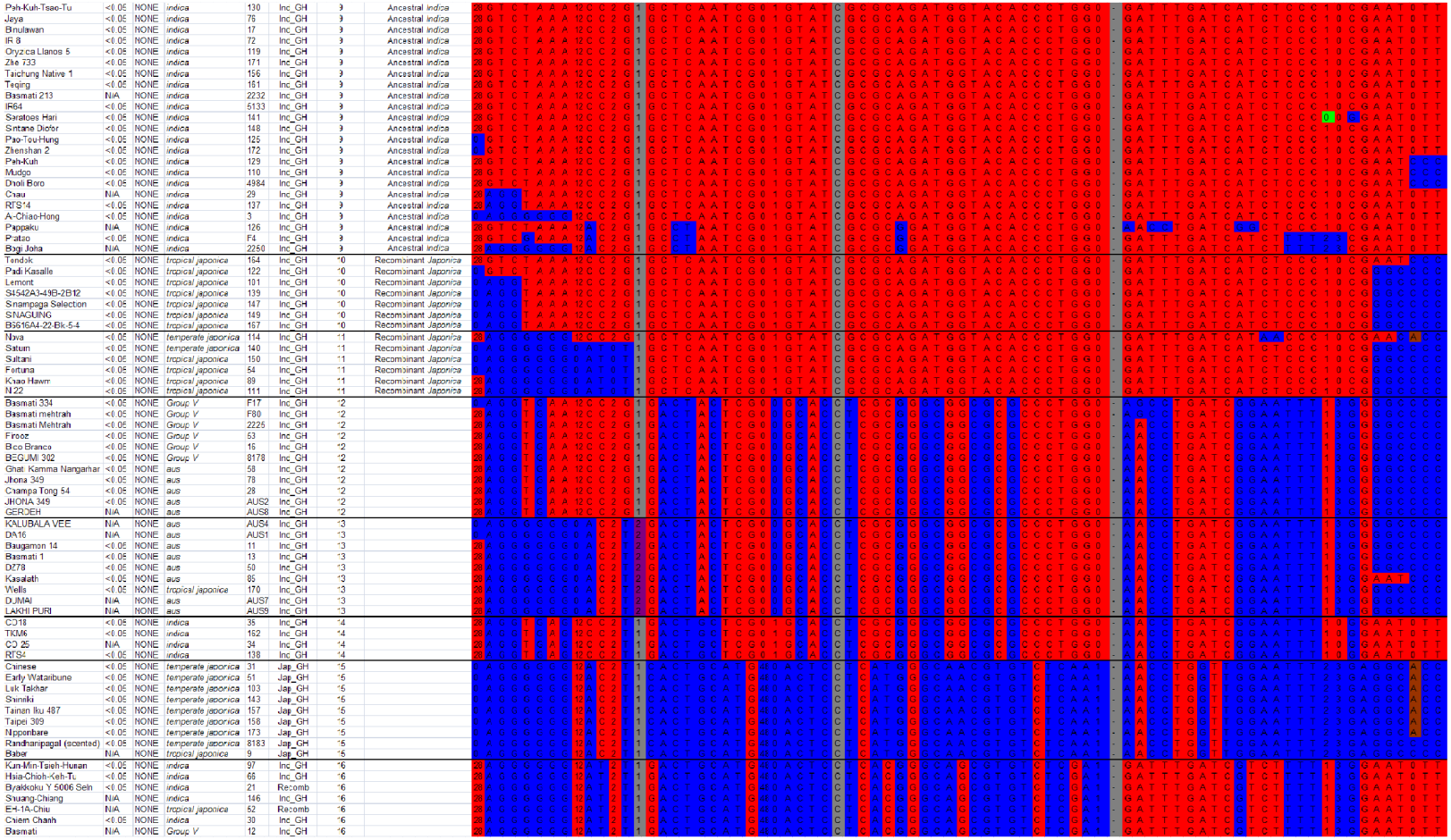
Extended Haplotypes for All 242 O. sativa Accessions and the Heterozygous Wild Accession

Accession Name	[ZAP]	FNP	Subpopulation	ID #	Gene Haplotype Cluster	Extended Haplotype Group #	Consensus Extended Haplotype	Polymorphisms * Position in Figure 1B
NORN 20	<0.05	NONE	*temperate japonica*	113	Jap_GH	1	Ancestral *Japonica*	
Hu Le Tao	<0.05	NONE	*temperate japonica*	67	Jap_GH	1	Ancestral *Japonica*	
GEUMOBYEO	<0.05	NONE	*temperate japonica*	56	Jap_GH	1	Ancestral *Japonica*	
Kamenoo	<0.05	NONE	*temperate japonica*	83	Jap_GH	1	Ancestral *Japonica*	
Jeuiku 393G	<0.05	NONE	*temperate japonica*	79	Jap_GH	1	Ancestral *Japonica*	
Koshihikari	<0.05	NONE	*temperate japonica*	94	Jap_GH	1	Ancestral *Japonica*	
Ta Mao Tsao	<0.05	NONE	*temperate japonica*	155	Jap_GH	1	Ancestral *Japonica*	
Ostiglia	<0.05	NONE	*temperate japonica*	121	Jap_GH	1	Ancestral *Japonica*	
NPE 836	<0.05	NONE	*temperate japonica*	115	Jap_GH	1	Ancestral *Japonica*	
Gotak Gatik	<0.05	NONE	*tropical japonica*	60	Jap_GH	1	Ancestral *Japonica*	
Rameo	<0.05	NONE	*temperate japonica*	134	Jap_GH	2	Ancestral *Japonica*	
Cs-M3	N/A	NONE	*temperate japonica*	36	Jap_GH	2	Ancestral *Japonica*	
Pato De Gallinazo	<0.05	NONE	*temperate japonica*	126	Jap_GH	2	Ancestral *Japonica*	
TA HUNG KU	<0.05	NONE	*temperate japonica*	154	Jap_GH	2	Ancestral *Japonica*	
Leung Pratew	<0.05	NONE	*temperate japonica*	102	Jap_GH	2	Ancestral *Japonica*	
Agostano	<0.05	NONE	*temperate japonica*	1	Jap_GH	2	Ancestral *Japonica*	
Oro	<0.05	NONE	*temperate japonica*	118	Jap_GH	2	Ancestral *Japonica*	
Davao	<0.05	NONE	*tropical japonica*	42	Jap_GH	2	Ancestral *Japonica*	
Be Ma Mua	N/A	NONE	*tropical japonica*	8395	Jap_GH	2	Ancestral *Japonica*	
L-202	<0.05	NONE	*tropical japonica*	98	Jap_GH	3	Ancestral *Japonica*	
DA 5	<0.05	NONE	*tropical japonica*	38	Jap_GH	3	Ancestral *Japonica*	
Kotobuki Mochi	<0.05	NONE	*tropical japonica*	95	Jap_GH	3	Ancestral *Japonica*	
Moroberekan	<0.05	NONE	*tropical japonica*	108	Jap_GH	3	Ancestral *Japonica*	
Canella De Ferro	<0.05	NONE	*tropical japonica*	23	Jap_GH	3	Ancestral *Japonica*	
Cnahora 144	<0.05	NONE	*tropical japonica*	27	Jap_GH	3	Ancestral *Japonica*	
Mirti	<0.05	NONE	*tropical japonica*	107	Jap_GH	3	Ancestral *Japonica*	
NPE 844	<0.05	NONE	*tropical japonica*	116	Jap_GH	3	Ancestral *Japonica*	
Knastano	<0.05	NONE	*tropical japonica*	92	Jap_GH	3	Ancestral *Japonica*	
IRAT 177	<0.05	NONE	*tropical japonica*	73	Jap_GH	3	Ancestral *Japonica*	
Honduras	<0.05	NONE	*tropical japonica*	65	Jap_GH	3	Ancestral *Japonica*	
OS6	<0.05	NONE	*tropical japonica*	120	Jap_GH	3	Ancestral *Japonica*	
Dourado Agulha	<0.05	NONE	*tropical japonica*	46	Jap_GH	3	Ancestral *Japonica*	
IAC 25	<0.05	NONE	*tropical japonica*	69	Jap_GH	3	Ancestral *Japonica*	
LAC 23	<0.05	NONE	*tropical japonica*	99	Jap_GH	3	Ancestral *Japonica*	
RF 1631-69	<0.05	NONE	*tropical japonica*	136	Jap_GH	3	Ancestral *Japonica*	
Ciba 65	<0.05	NONE	*tropical japonica*	37	Jap_GH	3	Ancestral *Japonica*	
Tiembese	<0.05	NONE	*tropical japonica*	166	Jap_GH	4	Ancestral *Japonica*	
Dammali	<0.05	NONE	*tropical japonica*	41	Jap_GH	4	Ancestral *Japonica*	
KJ115	N/A	NONE	*tropical japonica*	96	Jap_GH	4	Ancestral *Japonica*	
Farangey	N/A	NONE	*Group V*	8181	Jap_GH	4	Ancestral *Japonica*	
AAC 13523	N/A	NONE	*Group V*	8176	Jap_GH	4	Ancestral *Japonica*	
AAC 6070	<0.05	NONE	*Group V*	8177	Jap_GH	4	Ancestral *Japonica*	
AAC 13829	<0.05	NONE	*Group V*	F58	Jap_GH	4	Ancestral *Japonica*	
Basmati 370B	N/A	NONE	*Group V*	2237	Jap_GH	4	Ancestral *Japonica*	
Basmati 150	N/A	NONE	*Group V*	2239	Jap_GH	4	Ancestral *Japonica*	
Cahote Dhan	<0.05	NONE	*Group V*	F79	Jap_GH	4	Ancestral *Japonica*	
Tchampa	<0.05	NONE	*Group V*	160	Jap_GH	4	Ancestral *Japonica*	
P. Atok	N/A	NONE	*aus3 1 tropical japonica*	J152	Jap_GH	5	Fragrant *tropical japonica*	
P. Ikeng	N/A	NONE	*aus3 1 tropical japonica*	J187	Jap_GH	5	Fragrant *tropical japonica*	
P. Telengusan	N/A	NONE	*aus3 1 tropical japonica*	J178	Jap_GH	5	Fragrant *tropical japonica*	
Aias	1.37	NONE	*aus3 1 tropical japonica*	7	Jap_GH	5	Fragrant *tropical japonica*	
Bronyo	N/A	NONE	*aus3 1 tropical japonica*	15	Jap_GH	5	Fragrant *tropical japonica*	
P. Gadis	N/A	NONE	*aus3 1 tropical japonica*	J55	Jap_GH	5	Fragrant *tropical japonica*	
Aase Y Purg	0.42	NONE	*aus3 1 tropical japonica*	8	Jap_GH	5	Fragrant *tropical japonica*	
Dalla	1.43	NONE	*aus3 1 tropical japonica*	391	Jap_GH	5	Fragrant *tropical japonica*	
P. Pulut Longbanga	N/A	NONE	*aus3 1 tropical japonica*	J13	Jap_GH	5	Fragrant *tropical japonica*	
Azucena	0.31	NONE	*aus3 1 tropical japonica*	7737	Jap_GH	5	Fragrant *tropical japonica*	
Bdor	N/A	NONE	*aus3 1 tropical japonica*	8362	Jap_GH	5	Fragrant *tropical japonica*	
Careng	N/A	NONE	*aus3 1 tropical japonica*	8364	Jap_GH	5	Fragrant *tropical japonica*	
P. Pulut Mbau	N/A	NONE	*aus3 1 tropical japonica*	J176	Jap_GH	5	Fragrant *tropical japonica*	
WG 6	0.53	NONE	*aus3 1 tropical japonica*	159	Jap_GH	5	Fragrant *tropical japonica*	
Keriting Tingii	0.47	NONE	*aus3 1 tropical japonica*	87	Jap_GH	5	Fragrant *tropical japonica*	
Gogc Lempuk	N/A	NONE	*aus3 1 tropical japonica*	69	Jap_GH	5	Fragrant *tropical japonica*	
JC1	0.31	NONE	*aus3 1 Group V*	F3	Jap_GH	5	Fragrant Group V	
Ambemohar 1	N/A	NONE	*aus3 1 Group V*	2221	Jap_GH	5	Fragrant Group V	
BASFUL 714	0.30	NONE	*aus3 1 Group V*	8160	Jap_GH	5	Fragrant Group V	
JC220	0.47	NONE	*aus3 1 Group V*	8188	Jap_GH	5	Fragrant Group V	
AMBEMOHOR 157	0.34	NONE	*aus3 1 Group V*	8157	Jap_GH	5	Fragrant Group V	
Sadri Ghermes	N/A	NONE	*aus3 1 Group V*	8514	Jap_GH	5	Fragrant Group V	
Gerdeh	0.35	NONE	*aus3 1 Group V*	65	Jap_GH	5	Fragrant Group V	
Basmati bahar	0.27	NONE	*aus3 1 Group V*	2226	Jap_GH	5	Fragrant Group V	

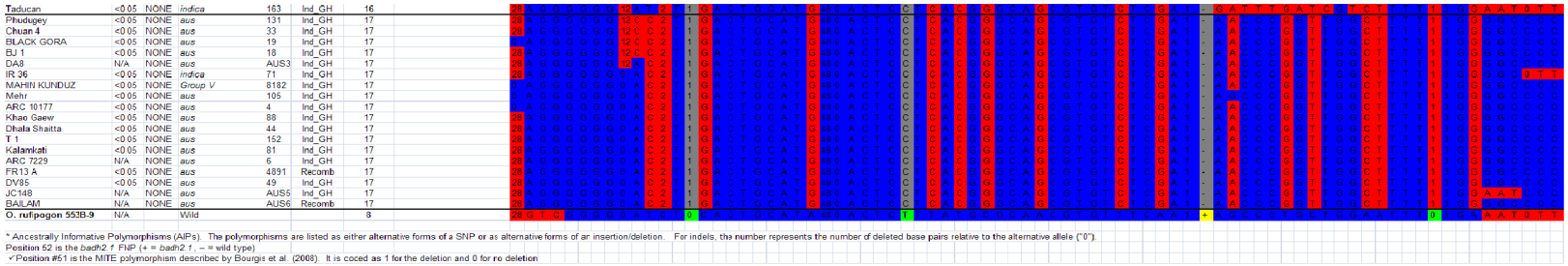
Supplemental Table 4.3 (continued)



Supplemental Table 4.3 (continued)



Supplemental Table 4.3 (continued)



Taducan	<0.05	NONE	*indica*	163	Ind_GH	16
Phudugey	<0.05	NONE	*aus*	131	Ind_GH	17
Chuan 4	<0.05	NONE	*aus*	33	Ind_GH	17
BLACK GORA	<0.05	NONE	*aus*	19	Ind_GH	17
BJ 1	<0.05	NONE	*aus*	18	Ind_GH	17
DA8	N/A	NONE	*aus*	AUS3	Ind_GH	17
IR 36	<0.05	NONE	*indica*	71	Ind_GH	17
MAHIN KUNDUZ	<0.05	NONE	*Group V*	8182	Ind_GH	17
Mehr	<0.05	NONE	*aus*	105	Ind_GH	17
ARC 10177	<0.05	NONE	*aus*	4	Ind_GH	17
Khao Gaew	<0.05	NONE	*aus*	88	Ind_GH	17
Dhala Shaitta	<0.05	NONE	*aus*	44	Ind_GH	17
T 1	<0.05	NONE	*aus*	152	Ind_GH	17
Kalamkati	<0.05	NONE	*aus*	81	Ind_GH	17
ARC 7229	N/A	NONE	*aus*	6	Ind_GH	17
FR13 A	<0.05	NONE	*aus*	4891	Recomb	17
DV85	<0.05	NONE	*aus*	49	Ind_GH	17
JC148	N/A	NONE	*aus*	AUS5	Ind_GH	17
BAILAM	N/A	NONE	*aus*	AUS6	Recomb	17
O. rufipogon 553B-9	N/A		Wild			8

* Ancestrally Informative Polymorphisms (AIPs). The polymorphisms are listed as either alternative forms of a SNP or as alternative forms of an insertion/deletion. For indels, the number represents the number of deleted base pairs relative to the alternative allele ("0").
Position 52 is the *badh2.1* FNP (+ = *badh2.1*, – = wild type)
✓Position #51 is the MITE polymorphism describec by Bourgis et al. (2008). It is coced as 1 for the deletion and 0 for no deletion

Supplemental Table 4.4: Novel Coding Mutations in *BADH2*

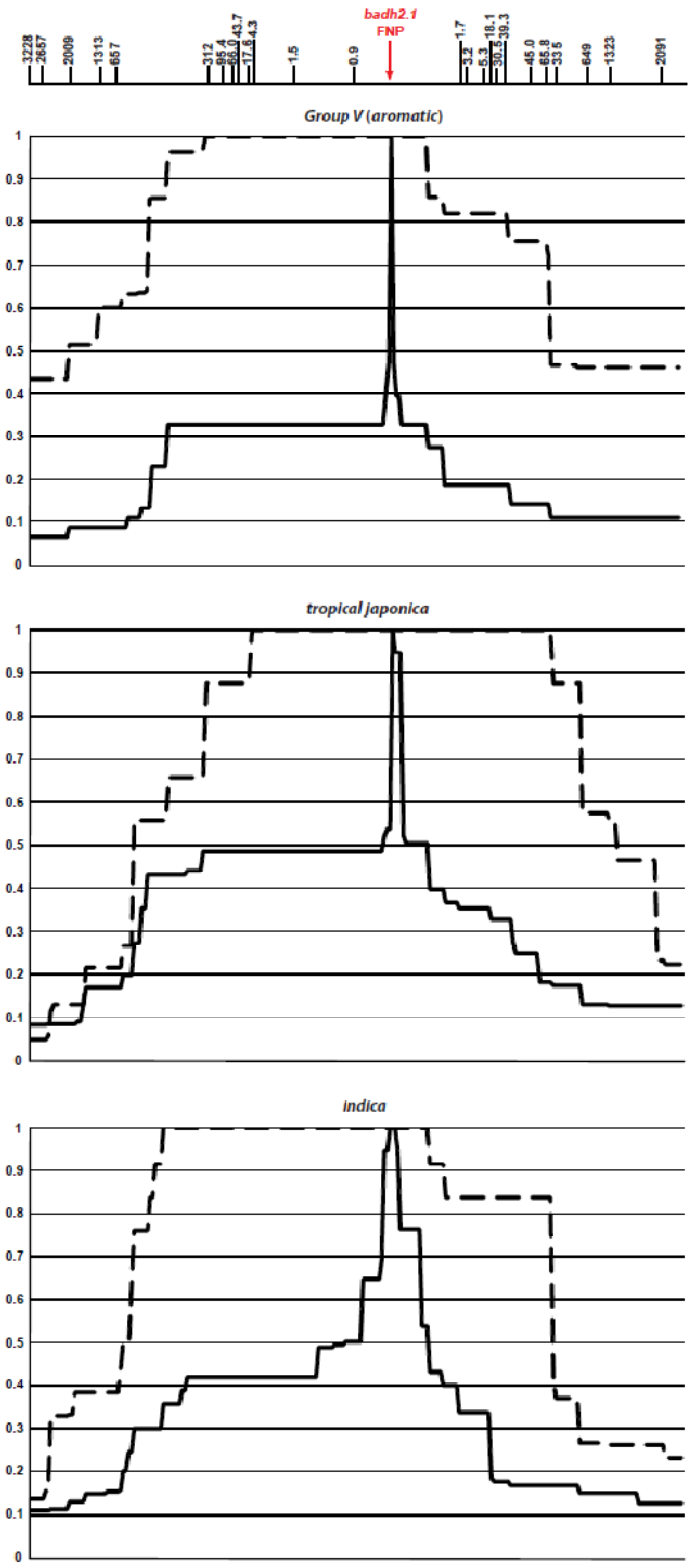
Allele #	IRGC #	Code #	Accession Name	Origin	Subpopulation	[2AP]	Marker	Mutation	Exon
badh2.2	28473	ANF4/8203	Hsiang-Keng-Nuo	China	*tropical japonica*	0.66	BadHapG1	7 bp deletion	2nd
badh2.3	69874	ANF10	Rasomotrafotsy	Madagascar	*tropical japonica*	0.74	BadHapG1	2 bp deletion	1st
badh2.4	27348	ANF3	Pare Baine Pulut	Indonesia	*tropical japonica*	0.59	BadHapG7	__ ➔T (1 bp insertion)	10th
badh2.5	52787	ANF9	Vishunparag	India	*aus*	0.41	BadHapG7	G➔T (substitution)	10th
badh2.6	13453	ANF1	Padi Wangi	Malaysia	*indica*	0.7	BadHapG7	T➔_ (1 bp deletion)	10th
badh2.7	34975	ANF6	Katak Tara	India	*aus*	0.37	BadHapG9	_➔G (1 bp insertion)	14th
	36394	ANF7	Suwanda Al	Sri Lanka	*aus*	0.3	BadHapG9		14th
	44204	ANF8	UPRB28	India	*aus*	0.42	BadHapG9		14th
	82408	ANF11	Basmati Bahar (Longer Grain)	India	*aus*	0.38	BadHapG9		14th
	83308	ANF12	Basmati	Nepal	*aus*	0.32	BadHapG9		14th
	83309	ANF13	Basmati	Nepal	*aus*	0.77	BadHapG9	__ insertion)	14th
badh2.8	33858	ANF5	Yangon Saba	Myanmar	Group V	0.31	BadHapG9		13th
	32960	F22	Balugyun	Myanmar	Group V	0.42	BadHapG9		13th
	97793	SR5	Paw San Hmwe	Myanmar	Group V	0.43	BadHapG9		13th
	N/A	8149	Paw San Hmwe (META-PHOR)	Myanmar	Group V	0.49	BadHapG9		13th
	33064	8172	Emata Longyu	Myanmar	Group V	0.14	BadHapG9		13th
badh2.9	18438	ANF2	Pandan Wangi	Indonesia	*tropical japonica*	0.29	BadHapG9	G➔T (substitution)	14th
	18438	PW1	Pandan Wangi	Indonesia	*tropical japonica*	0.19	BadHapG9		14th
	N/A	SR8	Pandan Wangi	Indonesia	*tropical japonica*	0.14	BadHapG9		14th
	N/A	SR9	Pandan Wangi	Indonesia	*tropical japonica*	0.18	BadHapG9		14th
	79569	8517	Mentik Wangi	Indonesia	*tropical japonica*	N/A	BadHapG9		14th
	N/A	8151	Pandan Wangi8 (META-PHOR)	Indonesia	*tropical japonica*	0.08	BadHapG9		14th
badh2.10	78746	8192	Tainung 72	Taiwan	*tropical japonica*	0.35	BadHapG9	C➔T (substitution)	13th
	5394	8215	Shiratama	Japan	*tropical japonica*	N/A	BadHapG9		13th
WT	90687	ANF14	Khau Tan Luong	Vietnam	*tropical japonica*	0.79			
	LG6732	K2/C7/8156	Kay Noi Leuang	Laos	*tropical japonica*	0.76			

Supplemental Table 4.5: Primers Used in This Study

Included in Fig. 1A	Marker Has AIP; Included in Fig. 1B	Included in Table 2 & Figure 3	Old Marker Name	New Marker Name	Start	End	Forward	Reverse	Tm
		X	BADHAPup9	BadhapUP12	17020003	17019198	CTCTACGTACGTCCACTTGATGA	GACCTGGTTTGACGGGAATA	55
		X	BADHAPup8	BadhapUP11	17589371	17590115	TGATCTTCAAAATGTTGCTTCC	TCGCCTTTTATAAGACCAGTCC	55
		X	BADHAPup7	BadhapUP10	18237475	18238153	AATGTGGGGCACAAGTAAATG	CCATTGACTTCGCAGTTCG	55
	X	X	ARO19	BadhapUP9	18933350	18933699	CCACCCTTTAGAAAGCCAAGT	GGACACATATCGGAGCGTATC	55
	X	X	BADHAPup6.5	BadhapUP8	19588765	19589793	CAAAATCGTAAAACGGGATGAG	CTTCTTAGCTGAAGGCTGAACG	55
		X	BADHAPup6	BadhapUP7	19934276	19935046	ATGGAACAGCACTTGGCATC	CACGATGGTGCTCCAGGAT	55
	X	X	BADHAPup4	BadhapUP6	20150703	20151430	CATTGGCATCCTCTACACCAT	CCACCAATGATCACTCTCTCTT	55
	X	X	BADHAPup3	BadhapUP5	20180116	20180845	GCCGGAGGTATGACATGGA	TCCTGACAACGGTCCAGATG	55
	X	X	BADHAPup2	BadhapUP4	20202769	20203215	TCCCCATTGTGGTGGTACA	CCGTCAAAGGTAATGGTCACT	55
	X	X	BADHAPup1	BadhapUP3	20228635	20229212	GAAGCAAGTGGAATTGCAAAA	GCAGTTGGCCACATAAACAA	55
	X	X	ARO9	BadhapUP2	20242081	20242544	CATGAATGTTCCCGTTGAAA	GCAGGTGGCAGTCCACTACT	55
	X	X	PRO2	BadhapUP1	20244602	20245371	CTTGGTGGTAAGTCGATGTCC	TTGTCTCTGCAATGAAGCTTGT	58
	X	X	ARO6	MITE	20245672	20245971	ACCGACTTAAATACGAACGATC	GTCAAACTCCCGACGTCATA	55
X		X	IN1	BadhapG1	20246844	20247918	CGAAGTCCGTACCAACTGC	GGCCGTGAGCCATATACACT	55
X		X	IN2	BadhapG2	20247795	20248554	AGTTGGAAGCATGGCTGATT	CCAGCTCAGATTTCCTCTCG	55
		X	IN3	BadhapG3	20248572	20248829	GATTGTGGGAAGCCTCTTGA	CGATAGGCTCTTTCCGAAGAT	55
X		X	IN4	BadhapG4	20248809	20249591	ATCTTCGGAAAGAGCCTATCG	AGGAGCTACCTTCCATGTTGC	55
X		X	IN5	BadhapG5	20249571	20250262	GCAACATGGAAGGTAGCTCCT	CCACCAAGTTCCAGTGAAACAG	55
X		X	IN6	BadhapG6	20250241	20251070	CTGTTTCACTGGAACTTGGTGG	GAATAAGACGCGATGTTGCACT	55
X		X	IN7	BadhapG7	20251049	20251510	AGTGCAACATCGCGTCTTATTC	CCCTCTTCAAGTGGATCTGACA	55
X		X	IN8	BadhapG8	20251489	20252254	TGTCAGATCCACTTGAAGAGGG	GAGTATCGTTGGCCAATTCAATG	55
X		X	IN9	BadhapG9	20252232	20252941	CATTGAATTGGCCAACGATACTC	GGCGTACTCCGTCACTTGCT	55
	X	X	END1	BadhapDOWN1	20254617	20255330	ATAGTGATTCAACGGCAGCAT	CGACATCTAGCGAGCATTTG	55
	X	X	END2	BadhapDOWN2	20256134	20256913	GCGGTTGATCTCTCGTACC	CAAATGGCAACTACCACCAT	55
	X	X	ARO1.5	BadhapDOWN3	20258275	20259051	AGGAAATGTGCGACGTCTGT	CGTGACCACCTAAGCCGTAT	55
	X	X	BADHAPdown1	BadhapDOWN4	20271039	20271629	TTGAAAGATGAGAACGGCAC	GAAATGCTACCTGAGGATTTGA	55
		X	BADHAPdown2	BadhapDOWN5	20283408	20283912	TTCGAGGCGTCATCAATTT	AAATGAGACCAGGAGTTCCAAT	55
	X	X	BADHAPdown2.5	BadhapDOWN6	20292211	20293044	GTTGTGCTCACACAGCTTGA	CAGATTATCCCACTCGAAATCA	55
	X	X	BADHAPdown3	BadhapDOWN7	20297924	20298476	AGGCCGAACTTCACGTTGT	CTTGGCCCCACCATTACAT	55
		X	BADHAPdown4	BadhapDOWN8	20318747	20319033	CCAGGAAAGCTGCTGCAC	GTCGTAGGAGTCGGCCTTG	55
	X	X	BADHAPdown6	BadhapDOWN9	20587910	20588319	CAATTGTTCAAGACGCACCA	AGTCGAGAATCCTCCATCTTGC	55
		X	BADHAPdown7	BadhapDOWN10	20901884	20902589	CTCCCTGAGGTGTTCTTGATG	TCTTGCTGAAACCTGGGTATG	55
	X	X	BADHAPdown8	BadhapDOWN11	21575778	21576441	GAATTTCGTGTGCCAGGCTA	CGGCGTTGACGACCTGTA	55
		X	BADHAPdown9	BadhapDOWN12	22344208	22345014	TCTTGCTGAAGGCGACCTAT	TTTCGCGTCTTTCTTGTGC	55

Supplemental Figure 4.1: Extended Haplotype Homozygosity (EHH) across the *BADH2* genomic region in individual subpopulations

EHH values were calculated for the *O. sativa* accessions examined in this study based on haplotype data across a 5.3 Mb genomic region surrounding the *BADH2* gene. In this figure, the EHH values were calculated for accessions within each subpopulation that possessed the *badh2.1* allele: *Group V* (n=67), *tropical japonica* (n=54), and *indica* (n=63). Solid and dashed lines indicate the combined EHH values of accessions having the wild type and *badh2.1* alleles, respectively. The position of the *badh2.1* FNP is indicated with an arrow. The locations of each amplicon used to obtain haplotype data are depicted across the top (Table S5) in terms of their physical distance from the *BADH2* gene (in kb). This figure demonstrates that the pattern of EHH decay around the *BADH2* gene for all of *O. sativa* (Figure 3) is very similar to the pattern observed within each subpopulation individually.


badh2.1
FNP
3228
26.57
2009
1313
657
312
85.4
80.0
43.7
17.6
4.3
1.5
0.9
1.7
5.2
5.3
18.1
30.5
39.3
45.0
65.8
33.5
649
1323
2091
Group V (aromatic)
tropical japonica
indica
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

REFERENCES

Ahn S.N., Bollich C.N., Tanksley S.D. (1992) RFLP tagging of a gene for aroma in rice. Theoretical and Applied Genetics 84:825-828

Amarawathi Y., Singh R., Singh A., Singh V., Mohapatra T., Sharma T., Singh N. (2008) Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). Molecular Breeding 21:49-65

Bashir K., Husnain T., Fatima T., Latif Z., Aks Mehdi S., Riazuddin S. (2004) Field evaluation and risk assessment of transgenic *indica* basmati rice. Molecular Breeding 13:301-312

Bhattacharjee P., Singhal R.S., Kulkarni P.R. (2002) Basmati rice: a review. International Journal of Food Science & Technology 37:1-12

Bourgis F., Guyot R., Gherbi H., Tailliez E., Amabile I., Salse J., Lorieux M., Delseny M., Ghesquière A. (2008) Characterization of the major fragance gene from an aromatic *japonica* rice and analysis of its diversity in Asian cultivated rice. Theoretical and Applied Genetics 117:353-368

Bradbury L., Gillies S., Brushett D., Waters D., Henry R. (2008) Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. Plant Molecular Biology 68:439-449

Bradbury L., Henry R., Jin Q., Reinke R.F., Waters D.L.E. (2005a) A perfect marker for fragrance genotyping in rice. Molecular Breeding 16:279-283

Bradbury L.M.T., Fitzgerald T.L., Henry R.J., Jin Q., Waters D.L.E. (2005b) The gene for fragrance in rice. Plant Biotechnology Journal 3:363-370

Buttery R.G., Ling L.C., Juliano B.O. (1982) 2-acetyl-1-pyrroline: An important aroma component of cooked rice. Chem Ind (London) 12:958-959

Buttery R.G., Ling L.C., Juliano B.O., Turnbaugh J.G. (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. Journal of Agricultural and Food Chemistry 31:823-826

Caicedo A.L., Williamson S.H., Hernandez R.D., Boyko A., Fledel-Alon A., York T.L., Polato N.R., Olsen K.M., Nielsen R., McCouch S.R., Bustamante C.D., Purugganan M.D. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genetics 3:e163

Chen L.J., Sun L.D., Song Z.P., Suh H.S., Lu B.R. (2004) Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Annals of Botany 93:67-73

Chen S., Wu J., Yang Y., Shi W., Xu M. (2006) The *fgr* gene responsible for rice fragrance was restricted within 69 kb. Plant Science 171:505-514

Chen S., Yang Y., Shi W., Ji Q., He F., Zhang Z., Cheng Z., Liu X., Xu M. (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. Plant Cell 20:1850-1861

Chou S.L. (1948) China is the place of origin of rice. J Rice Soc 7:53-54

Dellaporta S., Wood J., Hicks J. (1983) A plant DNA minipreparation: Version II. Plant Molecular Biology Reporter 1:19-21

Engle L.M., Chang T.T., Ramirez D.A. (1969) The cytogenetics of sterility in F1 hybrids of *Indica*/*Indica* and *Indica*/*Javanica* varieties of rice *Oryza sativa* L. Philipp Agric 53:289-307

Fitzgerald M.A., McCouch S.R., Hall R.D. (2009) Not just a grain of rice: the quest for quality. Trends in Plant Science 14:133-139

Fitzgerald M.A., Sackville Hamilton N.R., Calingacion M.N., Verhoeven H.A., Butardo V.M. (2008) Is there a second fragrance gene in rice? Plant Biotechnology Journal 6:416-423

Garg A.K., Kim J.-K., Owens T.G., Ranwala A.P., Choi Y.D., Kochian L.V., Wu R.J. (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences of the United States of America 99:15898-15903

Garris A.J., McCouch S.R., Kresovich S. (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165:759-769

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638

Glaszmann J.C. (1985) A varietal classification of Asian cultivated rice (*Oryza sativa* L.) based on isozyme polymorphism. In: Banta S.J. (ed) Rice Genetics I. International Rice Research Institute, Manila, Philippines, pp 83-90

Glaszmann J.C. (1987) Isozymes and classification of Asian rice varieties. Theoretical and Applied Genetics 74:21-30

Izawa T. (2008) The process of rice domestication: A new model based on recent data. Rice 1:127-134

Izawa T., Konishi S., Shomura A., Yano M. (2009) DNA changes tell us about rice domestication. Current Opinion in Plant Biology 12:185-192

Jain S., Jain R., McCouch S. (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. Theoretical and Applied Genetics 109:965-977

Jodon N.E. (1944) The inheritance of flower fragrance and other characters in rice. Journal of the American Society of Agronomy 36:844-848

Kadam B.S., Patankar V.K. (1938) Inheritance of aroma in rice. Chronica Botanica 4:32

Kang H.W., Cho Y.G., Yoon U.H., Eun M.Y. (1998) A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. Plant Molecular Biology Reporter 16:90-90

Khush G.S., Dela Cruz N. (2001) Developing Basmati rice with high yield potential. In: Duffy R. (ed) Specialty Rices of the World: Breeding, Production and Marketing. FAO and Science Publishers, Enfield, NH, pp 15-18

Kovach M.J., Sweeney M.T., McCouch S.R. (2007) New insights into the history of rice domestication. Trends in Genetics 23:578-587

Maqbool S.B., Husnain T., Riazuddin S., Masson L., Christou P. (1998) Effective control of yellow stem borer and rice leaf folder in transgenic rice *indica* varieties Basmati 370 and M 7 using the novel δ-endotoxin *cry2A Bacillus thuringiensis* gene. Molecular Breeding 4:501-507

McCouch S., CGSNL (2008) Gene nomenclature system for rice. Rice 1:72-84

Olsen K.M., Purugganan M.D. (2002) Molecular evidence on the origin and evolution of glutinous rice. Genetics 162:941-950

Paule C.M., Powers J.J. (1989) Sensory and chemical examination of aromatic and nonaromatic rices. Journal of Food Science 54:343-346

Prathepha P. (2009) The fragrance (*fgr*) gene in natural populations of wild rice (*Oryza rufipogon* Griff.). Genetic Resources and Crop Evolution 56:13-18

Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z.P., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837

Sang T., Ge S. (2007) The puzzle of rice domestication. Journal of Integrative Plant Biology 49:760-768

Shi W., Yang Y., Chen S., Xu M. (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. Molecular Breeding 22:185-192

Shobha R.N. (1992) Efforts to improve indigenous basmati rice. Rice India 1:9-15

Singh R.K., Singh U.S., Khush G.S. (2000) Aromatic Rices. Science Publishers, Enfield, NH

Song Z.P., Lu B.-R., Zhu Y.G., Chen J.K. (2003) Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657-665

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D., McCouch S.R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3:e133

Takano-Kai N., Jiang H., Kubo T., Sweeney M., Matsumoto T., Kanamori H., Padhukasahasram B., Bustamante C., Yoshimura A., Doi K., McCouch S. (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics 182:1323-1334

Ting Y. (1957) The origin and evolution of cultivated rice in China. Acta Agron Sinica 8:243-260

Vanavichit A. (2004) Discovering genes for rice grain aroma. Proceedings of the 1st International Conference on Rice for the Future, Bangkok, Thailand

Vanavichit A. (2007) Molecular diversity and evolution of aromatic rice in Thailand. International Training Workshop, The Conservation and Utilization of Tropical/Subtropical Plant Genetic Resources. Taiwan Agricultural Research Institute, Council of Agriculture (TARIC), Taiwan, pp 131-134

Vaughan D.A., Lu B.-R., Tomooka N. (2008) The evolving story of rice evolution. Plant Science 174:394-408

Xu K., Xu X., Fukao T., Canlas P., Maghirang-Rodriguez R., Heuer S., Ismail A.M., Bailey-Serres J., Ronald P.C., Mackill D.J. (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442:705-708

CHAPTER 5:
THE ORIGIN OF FRAGRANCE IN NERICA1

ABSTRACT

In this study, we investigated the cause and origin of fragrance in NERICA1, a fragrant rice inbred line developed from an interspecific cross between two non-fragrant parents. The genetic cause of fragrance in NERICA1 was found to be due to a previously reported mutation in the *BADH2* gene, the same allele responsible for the majority of modern fragrant rice varieties. Haplotype analysis around the *BADH2* gene in NERICA1, its parents, and 95 other varieties carrying the *badh2.1* allele identified the source of the *badh2.1* allele in NERICA1 was a fragrant *tropical japonica* variety, WAB638-1, which had been growing in the vicinity of the NERICA1 nursery during varietal development.

INTRODUCTION

African cultivated rice, *Oryza glaberrima*, was domesticated by local farmers along the Niger River long before Asian cultivated rice, *Oryza sativa*, was introduced into Africa (Vaughan et al., 2004; Sweeney and McCouch, 2007). *O. glaberrima* is well adapted to the harsh agro-ecological conditions that are characteristic of West Africa, and possesses tolerances to a wide range of biotic and abiotic stresses. Yet, *O. glaberrima* varieties are plagued by low yield potential and other traits that are unfavorable for modern rice production, such as susceptibility to lodging and shattering (Jones et al., 1997). This prompted the West Africa Rice Development Association (WARDA; now called Africa Rice Center) to create interspecific hybrids between *O. sativa* and *O. glaberrima*, in an effort to bring the beneficial tolerances of *O. glaberrima* into a high-yielding *O. sativa* genetic background. Over 300

interspecific BC_2 inbred lines were developed by WARDA by crossing an upland *tropical japonica* variety, WAB56-104 (recurrent parent), with the *O. glaberrima* variety CG14 (donor parent). These inbred lines are now collectively referred to as the New Rice for Africa or NERICA lines (Jones et al., 1997).

Eighteen upland varieties have been released from this effort to date, all of which exhibit moderate yields and stress tolerances under West African environmental conditions. Surprisingly, the line named NERICA1 also possessed fragrance in the grain, a characteristic that both the WAB56-104 and CG14 parents lacked. Based on conversations with WARDA breeders involved in the development of the NERICA lines, we learned that several fragrant rice varieties were planted in a field near the breeding nursery during the development of NERICA1 (personal communication, Mande Semon, WARDA). This prompted an investigation into the cause and origin of fragrance in NERICA1.

Fragrant rice varieties have been cultivated for centuries in many regions of Asia and the Middle East, where they are highly valued (Itani, 1993). These varieties, including the widely recognized Basmati and Jasmine types, are also very popular in Africa, Europe, Australia and the USA, allowing them to garner substantially higher export prices over their non-fragrant counterparts (Bhattacharjee et al., 2002; Qiu and Zhang, 2003). Breeding for fragrance has therefore become a major goal of African rice breeding programs.

A single recessive gene controlling fragrance in rice (*fgr*) was identified and fine mapped on the long arm of chromosome 8 (Kadam and Patankar, 1938; Jodon, 1944; Ahn et al., 1992; Lorieux et al., 1996; Chen et al., 2006). Sequence analysis identified

a complex mutation in a betaine aldehyde dehydrogenase gene, *BADH2*, that was perfectly associated with fragrance in the germplasm tested (Bradbury et al., 2005b). The functional nucleotide polymorphism (FNP) creating the *badh2.1* allele was described as a tandem array of 3 single nucleotide polymorphisms (SNPs) and an 8 base pair (bp) deletion in the 7^{th} exon of the gene that truncates the BADH2 protein (Bradbury et al., 2005b; Chen et al., 2008). While the biosynthetic pathway leading to fragrance is still being investigated, a nonfunctional BADH2 protein was shown to cause accumulation of AB-ald, a precursor of 2-acetyl-1-pyrroline (Chen et al., 2008), which is the major fragrant compound in Basmati and Jasmine rice (Buttery et al., 1982). While other nonfunctional *BADH2* alleles causing fragrance have been identified (Shi et al., 2008; Kovach et al., 2009), a large survey of diverse fragrant rice germplasm from Asia demonstrated that the *badh2.1* allele is the predominant allele responsible for elevated 2AP production, and thus fragrance, in rice (Fitzgerald et al., 2008).

When it is of interest to determine the origin of a useful allele in a set of germplasm, one can create haplotypes for the target gene. A haplotype is a series of polymorphisms at closely linked loci that tend to be inherited together, which can provide a signature of ancestry. Haplotypes are constructed based on molecular polymorphisms across a region (gene) of interest, and by comparing these haplotypes between accessions, one can trace the origin of the target allele among the genetically divergent subpopulations of rice. Haplotype analysis has been used to trace the origin of the Wx^b, *rc*, *sw5*, *GS3*, and *BADH2* alleles in *O. sativa* (Yamanaka et al., 2004; Sweeney et al., 2007; Kovach et al., 2009; Takano-Kai et al., 2009). The objective of this study was to use haplotype analysis to determine the origin of the fragrance allele in NERICA1.

MATERIALS AND METHODS

Plant materials

WAB56-104, CG14, NERICA1, and 95 other accessions were used for haplotype analysis, as indicated in Supplemental Table 5.1. All of these accessions were grown under greenhouse conditions at Cornell University (30^{o}C day/27^{o}C night; 60% humidity) and leaves were collected for DNA extraction.

Haplotype analysis

Haplotype analysis was performed according to the methods described previously (Kovach et al., 2009). In addition to CG14, WAB56-104, and NERICA1, 95 accessions known to carry the *badh2.1* allele were analyzed across the *BADH2* region of chromosome 8 to attempt to determine the ancestry of the *badh2.1* allele in NERICA1. A total of seven amplicons across the coding region of the *BADH2* gene and 19 amplicons in the flanking regions were sequenced, spanning 3.2 Mb upstream and 2.1 Mb downstream of *BADH2*, and resulting in over 14 kb of total aligned sequence. A complete list of the primers for the 26 amplicons that were sequenced for this study can be found in Supplemental Table 5.2. PCR was conducted using modified PCR protocols described previously (Garris et al., 2005) with an annealing temperature of 55°C. For sequencing, 5 µl of diluted PCR product was treated with 1.3 µl EXO/SAP (containing 3 units exonuclease I and 1 unit shrimp alkaline phosphatase) and incubated at 37°C for 45 minutes followed by 80°C for 15 minutes. Sequencing was performed on ABI Prism 3700/3100 DNA analyzers (Applied Biosystems, Foster City, CA) at the Cornell Life Sciences Core Laboratories Center. Sequences were aligned using the CodonCode Aligner program (CodonCode, Dedham, MA) and the ends of amplicons were trimmed to remove low quality sequences. Singletons and ambiguous sites were re-sequenced as necessary.

RESULTS AND DISCUSSION

When the two original parents of the NERICA lines, WAB56-104 and CG14, were assayed with the allele-specific marker for *badh2.1* (Bradbury et al., 2005a), neither was found to carry the *badh2.1* allele, consistent with their non-fragrant phenotype (Figure 5.1). However, the *badh2.1* allele was detected in NERICA1 and 95 other fragrant rice varieties, including Basmati 217 and Jasmine85. This suggested the *badh2.1* allele in NERICA1 must have originated from an undocumented parent, presumably due to an outcrossing event during the development or amplification of NERICA1.

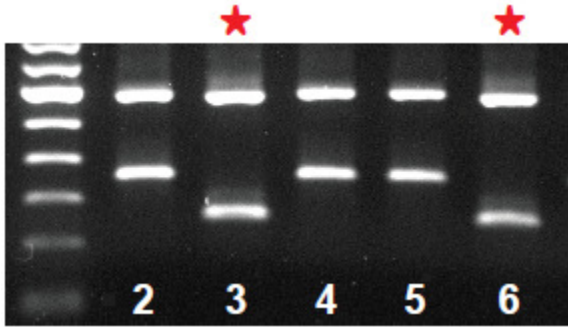


Figure 5.1: Results of the *badh2.1* allele-specific marker. The *badh2.1* allele-specific marker was amplified in (2) Nipponbare (wild-type control), (3) Basmati 217 (*badh2.1* control), (4) CG14, (5) WAB56-104, and (6) NERICA1. (1) is a DNA ladder. Lanes marked with a star are fragrant.

To determine the source of the fragrant allele in NERICA1, we commenced a "genetic paternity test" by performing haplotype analysis around the *BADH2* gene in NERICA1, WAB56-104, CG14, and 95 fragrant rice varieties, including several accessions known to have been grown at WARDA during the development of the NERICA lines. Our results demonstrated that NERICA1 had a 3.3Mb region flanking the *BADH2* gene that was 100% identical to the haplotype of a fragrant *tropical japonica* variety developed at WARDA, WAB638-1 (Figure 5.2). No other fragrant accession tested in this study matched the NERICA1 haplotype across this entire genomic region (Supplemental Table 5.3). These results demonstrate that the *badh2.1* allele in NERICA1 originated from WAB638-1.This agrees with the historical account given by WARDA breeders that WAB638-1 had flowered in a field near the breeding nursery during the development of the NERICA lines. NERICA1 appears to have experience several rounds of introgression from unknown paternal sources, as evidenced by the fact that the flanking regions around the introgression from WAB638-1 do not match either of the two known parents, but are similar to several *indica* and *tropical japonica* accessions from our germplasm panel (Figure 5.2; Supplemental Table 5.3). These results are also consistent with the findings of a molecular profiling experiment that discovered a large number of non-parental alleles in the NERICA lines (Semagn et al., 2007), and highlights the necessity of taking precautions to prevent outcrossing and contamination during inbred development and varietal release (Ikeda et al., 2007). At the same time, the introduction of fragrance into NERICA1 bestowed it with a grain quality trait that is highly preferred by African consumers, demonstrating that the injection of new genetic diversity during varietal development may be a more valuable alternative to simply generating "pure lines".

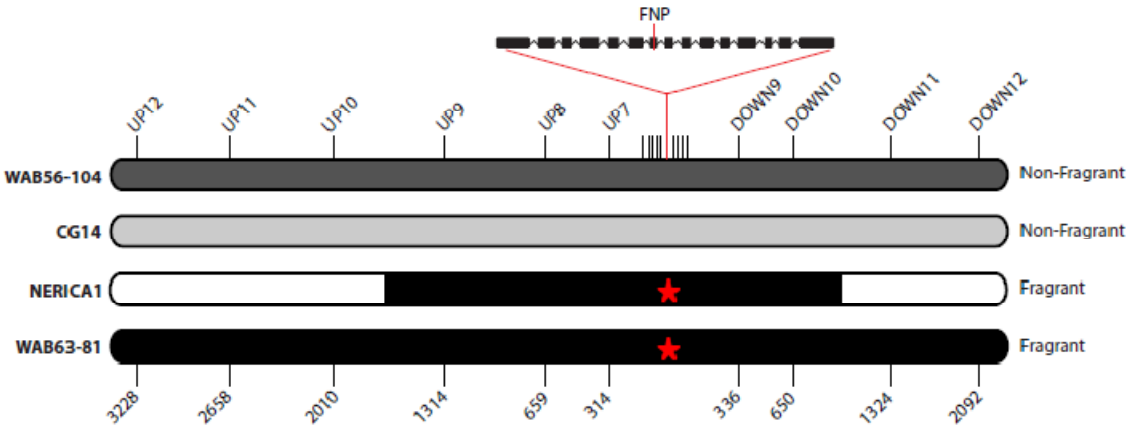


Figure 5.2: The origin of *badh2.1* in NERICA1. Nineteen regions flanking the *BADH2* gene were sample-sequenced in WAB56-104, CG14, NERICA1, and 95 other accessions that carry the *badh2.1* allele. All polymorphisms detected at a frequency above 5% in this population were concatenated to create extended haplotypes across the *BADH2* region. Each horizontal bar depicts the target region of chromosome 8, colored dark grey (WAB56-104 haplotype), light grey (CG14 haplotype), or black (WAB638-1 haplotype) with the fragrance phenotypes to the right of each bar. Breaks in coloration indicate positions where recombination was detected in NERICA1, with the lack of color indicating a haplotype of unknown origin. The position of the *BADH2* gene is indicated. The star indicates the presence of the *badh2.1* allele.

APPENDIX

Supplemental Table 5.1: Rice Accessions Used in This Study

A. Accessions used for Haplotype Analysis

ID	RA #	IRGC #	Accession Name	Origin	Subpopulation	*BADH2* allele
control	RA 7309	IRGC 117274	Nipponbare	Japan	*temperate japonica*	wild-type
control	RA 7150	IRGC 53637	Basmati 217	India	*Group V (aromatic)*	*badh2.1*
parent	RA 7374		WAB56-104		*tropical japonica*	wild-type
parent	RA 8609		CG14		*O. glaberrima*	wild-type
parent	RA 6111		NERICA1	Ghana	*O. sativa x O. glaberrima*	*badh2.1*
parent	RA 6130		IDSA 85	Ghana	*indica*	wild-type
Haplo1	RA 7141	IRGC 12440	ARC 10352	India	*Group V (aromatic)*	*badh2.1*
Haplo2	RA 7143	IRGC 43325	Arias	Indonesia	*tropical japonica*	*badh2.1*
Haplo3	RA 7144	IRGC 6949	Asse Y Pung	Philippines	*tropical japonica*	*badh2.1*
Haplo4	RA 7150	IRGC 53637	Basmati 217	India	*Group V (aromatic)*	*badh2.1*
Haplo5	RA 7151	IRGC 55457	Beonjo	Korea	*tropical japonica*	*badh2.1*
Haplo6	RA 7181	PI 584607	DOM-SOFID	Iran	*Group V (aromatic)*	*badh2.1*
Haplo7	RA 7191	IRGC 32300	Gerdeh	Iran	*Group V (aromatic)*	*badh2.1*
Haplo8	RA 7195	IRGC 43394	Gogo Lempuk	Indonesia	*tropical japonica*	*badh2.1*
Haplo9	RA 7223	IRGC 19972	Keriting Tingii	Indonesia	*tropical japonica*	*badh2.1*
Haplo10	RA 7229	IRGC 12793	Kitrana 508	Madagascar	*Group V (aromatic)*	*badh2.1*
Haplo11	RA 7248	IRGC 6298	N12	India	*Group V (aromatic)*	*badh2.1*
Haplo12	RA 7268	IRGC 8952	Rathuwee	Sri Lanka	*indica*	*badh2.1*
Haplo13	RA 7272	IRGC 8234	RTS12	Vietnam	*indica*	*badh2.1*
Haplo14	RA 7305	Clor 5309	WC 6	China	*tropical japonica*	*badh2.1*
Haplo15	RA 7527	Clor 9483	Della	USA	*tropical japonica*	*badh2.1*
Haplo16	RA 2221	IRGC 52266	Ambemohar 1	India	*Group V (aromatic)*	*badh2.1*
Haplo17	RA 2226	IRGC 79560	Basmati Bahar (shorter grain)	India	*Group V (aromatic)*	*badh2.1*
Haplo18	RA 2248	IRGC 53827	BPT1235	India	*Group V (aromatic)*	*badh2.1*
Haplo19	RA 2253		Barhail	India	*Group V (aromatic)*	*badh2.1*
Haplo20	RA 2254		Blomberg	India	*Group V (aromatic)*	*badh2.1*
Haplo21	RA 2258	IRGC 45298	Chinisakkar	India	*Group V (aromatic)*	*badh2.1*
Haplo22	RA 2259	IRGC 43805	Chinigura	Bangladesh	*Group V (aromatic)*	*badh2.1*
Haplo23	RA 2262	IRGC 78476	Dulhamiya	India	*Group V (aromatic)*	*badh2.1*
Haplo24	RA 2264	IRGC 8628	DD66	Bangladesh	*Group V (aromatic)*	*badh2.1*
Haplo25	RA 2271		Haryana Basmati1	India	*Group V (aromatic)*	*badh2.1*
Haplo26	RA 2277		HKR93-401	India	*Group V (aromatic)*	*badh2.1*
Haplo27	RA 2278		Tararori Basmati	India	*Group V (aromatic)*	*badh2.1*
Haplo28	RA 2279	IRGC 78422	Pusa basmati-1	India	*Group V (aromatic)*	*badh2.1*
Haplo29	RA 2280		HBC-19	India	*Group V (aromatic)*	*badh2.1*
Haplo30	RA 4878	IRGC 27748	Khao Dawk Mali 105	Thailand	*indica*	*badh2.1*
Haplo31	RA 7310	IRIS 2254730	Azucena	Philippines	*tropical japonica*	*badh2.1*
Haplo32	RA 8126		Q74	Malaysia	*indica*	*badh2.1*
Haplo33	RA 8127		Nang Thom Cho Dao	Vietnam	*indica*	*badh2.1*
Haplo34	RA 8130		RD33	Thailand	*indica*	*badh2.1*
Haplo35	RA 8131		PTT1	Thailand	*indica*	*badh2.1*
Haplo36	RA 8132		RD6	Thailand	*indica*	*badh2.1*
Haplo37	RA 8135		OM3536	Vietnam	*indica*	*badh2.1*
Haplo38	RA 8136		Basmati 2000	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo39	RA 8141		Nho Thom	Vietnam	*indica*	*badh2.1*
Haplo40	RA 8147		Chao do	Laos	*indica*	*badh2.1*
Haplo41	RA 8155		HNN	Laos	*indica*	*badh2.1*
Haplo42	RA 8157	IRGC 6148	AMBEMOHOR 157	India	*Group V (aromatic)*	*badh2.1*
Haplo43	RA 8158	IRGC 20310	ARC 6011	India	*Group V (aromatic)*	*badh2.1*
Haplo44	RA 8159	IRGC 27781	BARA 413	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo45	RA 8160	IRGC 29258	BASFUL 714	Bangladesh	*Group V (aromatic)*	*badh2.1*
Haplo46	RA 8162	IRGC 10330	BASMATI 6311	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo47	RA 8163	IRGC 9027	BASMATI C 622	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo48	RA 8169	IRGC 5857	DA13	Bangladesh	*Group V (aromatic)*	*badh2.1*
Haplo49	RA 8171	IRGC 32291	DOMSIAH	Iran	*Group V (aromatic)*	*badh2.1*
Haplo50	RA 8174	IRGC 27953	HARANDI 379	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo51	RA 8175	IRGC 32329	SADRI	Iran	*Group V (aromatic)*	*badh2.1*
Haplo52	RA 8185	IRGC 9026	BASMATI 370	India	*Group V (aromatic)*	*badh2.1*
Haplo53	RA 8186	IRGC 32985	BOKEHMWE	Myanmar	*indica*	*badh2.1*
Haplo54	RA 8187	IRGC 16107	GUOR LAO	Vietnam	*indica*	*badh2.1*
Haplo55	RA 8188	IRGC 9092	JC220	India	*Group V (aromatic)*	*badh2.1*
Haplo56	RA 8189	IRGC 33357	MA WAINE OHN	Myanmar	*indica*	*badh2.1*
Haplo57	RA 8191	IRGC 33571	PAWSANHMWE	Myanmar	*indica*	*badh2.1*
Haplo58	RA 8193	IRGC 78747	TAINUNG SEN 20	Taiwan	*indica*	*badh2.1*
Haplo59	RA 8195	IRGC 233103	Aus Basmati	India	*Group V (aromatic)*	*badh2.1*
Haplo60	RA 8200	IRGC 385453	Basmati 410	India	*Group V (aromatic)*	*badh2.1*
Haplo61	RA 8201	IRGC 385785	Basmati Lal	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo62	RA 8206	IRGC 412772	Basmati 802	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo63	RA 8208	IRGC 412852	Karnal Basmati	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo64	RA 8222	IRGC 584591	YUN AROMATIC GLUTINOUS	China	*indica*	*badh2.1*
Haplo65	RA 8224	IRGC 595927	JASMINE 85	USA	*indica*	*badh2.1*
Haplo66	RA 8230	GSOR 310319	BC5-55	India	*indica*	*badh2.1*
Haplo67	RA 8231	GSOR 311134	Sathi Basmati	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo68	RA 8232	GSOR 311216	Basmati 802	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo69	RA 8362	IRGC 77952	Bidor	Malaysia	*tropical japonica*	*badh2.1*
Haplo70	RA 8364	IRGC 77954	Coreng	Malaysia	*tropical japonica*	*badh2.1*
Haplo71	RA 8503	IRGC 9877	Kalijira	Bangladesh	*Group V (aromatic)*	*badh2.1*
Haplo72	RA 8514	IRGC 66247	Sadri Ghermes	Iran	*Group V (aromatic)*	*badh2.1*
Haplo73	RA 8516	IRGC 78480	Jasmine Scented	Thailand	*indica*	*badh2.1*
Haplo74	RA 8521	IRGC 87902	Somaly Krar-oob	Cambodia	*indica*	*badh2.1*
Haplo75			BAS270			*badh2.1*
Haplo76			BASPUNJAB			*badh2.1*
Haplo77	F14	IRGC 27798	Basmati 1_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo78	F16	IRGC 27815	Basmati 213_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo79	F18	IRGC 27822	Basmati 370B_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo80	F19	IRGC 27824	Basmati 372A_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo81	F20	IRGC 27869	Chahora 144_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo82	F21	IRGC 30238	Champa Tong 54_IRRI	Thailand	*indica*	*badh2.1*
Haplo83	F3	IRGC 9091	JC1_IRRI	India	*Group V (aromatic)*	*badh2.1*
Haplo84	F33	IRGC 78470	Basmati sathi_IRRI	India	*Group V (aromatic)*	*badh2.1*

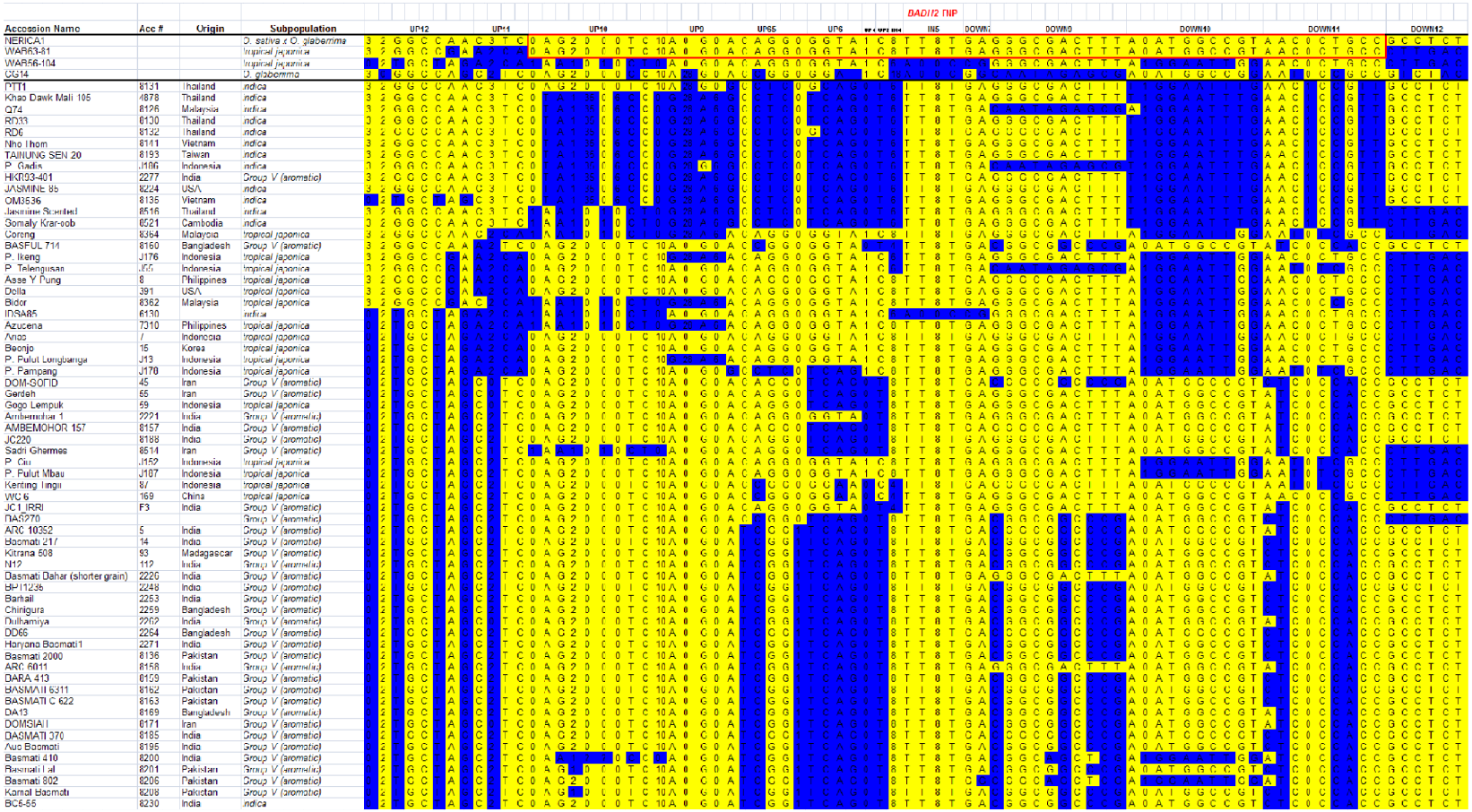
Supplemental Table 5.1 (continued)

Haplo85	F38	IRGC 5999	Pankhari 203_IRRI	India	*Group V (aromatic)*	*badh2.1*
Haplo86	F61	IRGC 9036	Basmati 6113_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo87	F63	IRGC 9030	Basmati 5836_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo88	F75	IRGC 27805	Basmati 122_IRRI	Pakistan	*Group V (aromatic)*	*badh2.1*
Haplo89	RA 5756	collected in Indo	P. Pulut Longbanga	Indonesia	*tropical japonica*	*badh2.1*
Haplo90	RA 5895	collected in Indo	P. Telengusan	Indonesia	*tropical japonica*	*badh2.1*
Haplo91	RA 5919	collected in Indo	P. Ciu	Indonesia	*tropical japonica*	*badh2.1*
Haplo92	RA 5921	collected in Indo	P. Ikeng	Indonesia	*tropical japonica*	*badh2.1*
Haplo93	RA 5929	collected in Indo	P. Pampang	Indonesia	*indica*	*badh2.1*
Haplo94	RA 5930	collected in Indo	P. Gadis	Indonesia	*tropical japonica*	*badh2.1*
Haplo95	RA 5798	collected in Indo	P. Pulut Mbau	Indonesia	*tropical japonica*	*badh2.1*
B. Accessions used for Association Analysis						
ID	**RA #**	**Accession Name**	**Proximal Origin**	**Species**	***BADH2* allele**	
Assoc1	RA 6112	Basmati Super	Ghana	*O. sativa*	wild-type	
Assoc2	RA 6113	Basmati Okyereko	Ghana	*O. sativa*	*badh2.1*	
Assoc3	RA 6114	Basmati 6129	Ghana	*O. sativa*	*badh2.1*	
Assoc4	RA 6116	Basmati Pusa	Ghana	*O. sativa*	*badh2.1*	
Assoc5	RA 6117	KCR Baika	Ghana	*O. sativa*	*badh2.1*	
Assoc6	RA 6118	Aromatic Short	Ghana	*O. sativa*	*badh2.1*	
Assoc7	RA 6119	Aromatic Long	Ghana	*O. sativa*	*badh2.1*	
Assoc8	RA 6120	Marshal	Ghana	*O. sativa*	*badh2.1*	
Assoc9	RA 6122	Cambodia	Ghana	*O. sativa*	*badh2.1*	
Assoc10	RA 6123	P38	Ghana	*O. sativa*	*badh2.1*	
Assoc11	RA 6124	IET 6279	Ghana	*O. sativa*	wild-type	
Assoc12	RA 6125	NERICA2	Ghana	*O. sativa x O. glaberrima*	wild-type	
Assoc13	RA 6126	NERICA3	Ghana	*O. sativa x O. glaberrima*	wild-type	
Assoc14	RA 6127	NERICA4	Ghana	*O. sativa x O. glaberrima*	wild-type	
Assoc15	RA 6128	NERICA5	Ghana	*O. sativa x O. glaberrima*	wild-type	
Assoc16	RA 6129	NERICA6	Ghana	*O. sativa x O. glaberrima*	wild-type	
Assoc17	RA 6131	Tox 3377	Ghana	*O. sativa*	wild-type	
Assoc18	RA 6132	Bouake 189	Ghana	*O. sativa*	wild-type	
Assoc19	RA 6133	Wita 7	Ghana	*O. sativa*	wild-type	
Assoc20	RA 6134	Tox 3108-56-4-2-2-2	Ghana	*O. sativa*	wild-type	
Assoc21	RA 6135	ITA 320	Ghana	*O. sativa*	wild-type	
Assoc22	RA 6136	ITA 324	Ghana	*O. sativa*	wild-type	
Assoc23	RA 6137	Digan	Ghana	*O. sativa*	wild-type	
Assoc24	RA 6138	Mui Wong	Ghana	*O. glaberrima*	wild-type	
Assoc25	RA 6139	Mui Gelik	Ghana	*O. glaberrima*	wild-type	
Assoc26	RA 6140	Boning Kari	Ghana	*O. glaberrima*	wild-type	
Assoc27	RA 6141	GH 1835	Ghana	*O. glaberrima*	wild-type	
Assoc28	RA 6142	GH 2497	Ghana	*O. glaberrima*	N/A	
Assoc29	RA 6143	GH 1796	Ghana	*O. glaberrima*	wild-type	
Assoc30	RA 6144	GH 1825	Ghana	*O. glaberrima*	N/A	
Assoc31	RA 6145	GH 1572	Ghana	*O. sativa*	wild-type	
Assoc32	RA 6146	GH 1527	Ghana	*O. sativa*	N/A	
Assoc33	RA 6147	GH 4006	Ghana	*O. sativa*	wild-type	
Assoc34	RA 6148	GH 4008	Ghana	*O. sativa*	wild-type	
Assoc35	RA 6149	GH 1533	Ghana	*O. sativa*	wild-type	
Assoc36	RA 6150	GH 1809	Ghana	*O. sativa*	N/A	
Assoc37	RA 6151	GH 1517	Ghana	*O. sativa*	*badh2.1*	
Assoc38	RA 6152	GH 1586	Ghana	*O. sativa*	wild-type	
Assoc39	RA 6153	GH 1584	Ghana	*O. sativa*	wild-type	
Assoc40	RA 6154	GH 1788	Ghana	*O. sativa*	N/A	
Assoc41	RA 6155	GH 1580	Ghana	*O. sativa*	wild-type	
Assoc42	RA 6156	GH 1521	Ghana	*O. sativa*	wild-type	
Assoc43	RA 6157	GH 1512	Ghana	*O. sativa*	wild-type	
Assoc44	RA 6158	GH 4012	Ghana	*O. sativa*	*badh2.1*	
Assoc45	RA 6159	GH 1585	Ghana	*O. sativa*	wild-type	
Assoc46	RA 6160	GH 1822	Ghana	*O. sativa*	N/A	
Assoc47	RA 6161	GH 1581	Ghana	*O. sativa*	wild-type	
Assoc48	RA 6162	GH 1574 (PR 5)	Ghana	*O. sativa*	wild-type	
Assoc49	RA 6163	GH 1575	Ghana	*O. sativa*	wild-type	
Assoc50	RA 6164	GH 1518	Ghana	*O. sativa*	wild-type	
Assoc51	RA 6165	GH 1593	Ghana	*O. sativa*	wild-type	
Assoc52	RA 6166	GH 1518	Ghana	*O. sativa*	N/A	
Assoc53	RA 6167	GH 1569	Ghana	*O. sativa*	*badh2.1*	
Assoc54	RA 6168	GH 1834	Ghana	*O. sativa*	wild-type	
Assoc55	RA 6169	GH 1801	Ghana	*O. sativa*	wild-type	
Assoc56	RA 6170	GH 1837	Ghana	*O. sativa*	wild-type	
Assoc57	RA 6171	GH 1513	Ghana	*O. sativa*	wild-type	
Assoc58	RA 6172	GH 1528	Ghana	*O. sativa*	wild-type	
Assoc59	RA 6173	GH 1570	Ghana	*O. sativa*	wild-type	
Assoc60	RA 6174	GH 1577	Ghana	*O. sativa*	wild-type	
Assoc61	RA 6175	GH 1536	Ghana	*O. sativa*	N/A	
Assoc62	RA 6176	GH 1582	Ghana	*O. sativa*	wild-type	
Assoc63	RA 6177	GH 1576	Ghana	*O. sativa*	wild-type	
Assoc64	RA 6178	GH 1511	Ghana	*O. sativa*	wild-type	
Assoc65	RA 6179	GH 1545	Ghana	*O. sativa*	wild-type	
Assoc66	RA 6180	GH 1524	Ghana	*O. sativa*	wild-type	
Assoc67	RA 6181	GH 1520	Ghana	*O. sativa*	wild-type	
Assoc68	RA 6182	GH 1790	Ghana	*O. sativa*	N/A	
Assoc69	RA 6183	GH 1546	Ghana	*O. sativa*	N/A	
Assoc70	RA 6184	GH 1589	Ghana	*O. sativa*	wild-type	
Assoc71	RA 6185	GH 1571	Ghana	*O. sativa*	wild-type	
Assoc72	RA 6186	GH 1590	Ghana	*O. sativa*	wild-type	
Assoc73	RA 6187	GH 1811	Ghana	*O. sativa*	N/A	
Assoc74	RA 6188	GH 1804	Ghana	*O. sativa*	N/A	
Assoc75	RA 6189	GH 1574	Ghana	*O. sativa*	N/A	
Assoc76	RA 6190	Viwonor (Tall)	Ghana	*O. glaberrima*	wild-type	
Assoc77	RA 6191	Viwonor (Short)	Ghana	*O. glaberrima*	wild-type	
Assoc78	RA 6192	Wiwotor	Ghana	*O. glaberrima*	wild-type	

Supplemental Table 5.2: Primers Used in This Study

Marker Name	Start	End	Forward	Reverse	Tm
BadhapUP12	17020003	17019198	CTCTACGTACGTCCACTTGATGA	GACCTGGTTTGACGGGAATA	55
BadhapUP11	17589371	17590115	TGATCTTCAAAATGTTGCTTCC	TCGCCTTTTATAAGACCAGTCC	55
BadhapUP10	18237475	18238153	AATGTGGGGCACAAGTAAATG	CCATTGACTTCGCAGTTCG	55
BadhapUP9	18933350	18933699	CCACCCTTTAGAAAGCCAAGT	GGACACATATCGGAGCGTATC	55
BadhapUP8	19588765	19589793	CAAAATCGTAAAACGGGATGAG	CTTCTTAGCTGAAGGCTGAACG	55
BadhapUP7	19934276	19935046	ATGGAACAGCACTTGGCATC	CACGATGGTGCTCCAGGAT	55
BadhapUP6	20150703	20151430	CATTGGCATCCTCTACACCAT	CCACCAATGATCACTCTCTCTT	55
BadhapUP5	20180146	20180845	GCCGGAGGTATGACATGGA	TCCTGACAACGGTCCAGATG	55
BadhapUP4	20202769	20203215	TCCCCATTGTGGTGGTACA	CCGTCAAAGGTAATGGTCACT	55
BadhapUP3	20228635	20229212	GAAGCAAGTGGAATTGCAAAA	GCAGTTGGCCACATAAACAA	55
BadhapUP2	20242081	20242544	CATGAATGTTCCCGTTGAAA	GCAGGTGGCAGTCCACTACT	55
BadhapG2	20247795	20248554	AGTTGGAAGCATGGCTGATT	CCAGCTCAGATTTCCTCTCG	55
BadhapG4	20248809	20249591	ATCTTCGGAAAGAGCCTATCG	AGGAGCTACCTTCCATGTTGC	55
BadhapG5	20249571	20250262	GCAACATGGAAGGTAGCTCCT	CCACCAAGTTCCAGTGAAACAG	55
BadhapG6	20250241	20251070	CTGTTTCACTGGAACTTGGTGG	GAATAAGACGCGATGTTGCACT	55
BadhapG7	20251049	20251510	AGTGCAACATCGCGTCTTATTC	CCCTCTTCAAGTGGATCTGACA	55
BadhapG8	20251489	20252254	TGTCAGATCCACTTGAAGAGGG	GAGTATCGTTGGCCAATTCAATG	55
BadhapG9	20252232	20252941	CATTGAATTGGCCAACGATACTC	GGCGTACTCCGTCACTTGCT	55
BadhapDOWN3	20258275	20259051	AGGAAATGTGCGACGTCTGT	CGTGACCACCTAAGCCGTAT	55
BadhapDOWN4	20271039	20271629	TTGAAAGATGAGAACGGCAC	GAAATGCTACCTGAGGATTTGA	55
BadhapDOWN5	20283408	20283912	TTCGAGGCGTCATCAATTT	AAATGAGACCAGGAGTTCCAAT	55
BadhapDOWN7	20297924	20298476	AGGCCGAACTTCACGTTGT	CTTGGCCCCACCATTACAT	55
BadhapDOWN9	20587910	20588319	CAATTGTTCAAGACGCACCA	AGTCGAGAATCCTCCATCTTGC	55
BadhapDOWN10	20901884	20902589	CTCCCTGAGGTGTTCTTGATG	TCTTGCTGAAACCTGGGTATG	55
BadhapDOWN11	21575778	21576441	GAATTTCGTGTGCCAGGCTA	CGGCGTTGACGACCTGTA	55
BadhapDOWN12	22344208	22345014	TCTTGCTGAAGGCGACCTAT	TTTCGCGTCTTTCTTGTGC	55

Supplemental Table 5.3: Haplotypes for 3.3Mb region Flanking *BADH2*



Supplemental Table 5.3 (continued)

Sathi Basmati	3231	Pakistan	*Group V (aromatic)*
Kaljira	3503	Bangladesh	*Group V (aromatic)*
Basmati 1_IRRI	F14	Pakistan	*Group V (aromatic)*
Basmati 213_IRRI	F16	Pakistan	*Group V (aromatic)*
Basmati 370B_IRRI	F18	Pakistan	*Group V (aromatic)*
Basmati 372A_IRRI	F19	Pakistan	*Group V (aromatic)*
Chahora 144_IRRI	F20	Pakistan	*Group V (aromatic)*
Basmati sathi_IRRI	F33	India	*Group V (aromatic)*
Parkhar 203_IRRI	F38	India	*Group V (aromatic)*
Basmati 6113_IRRI	F61	Pakistan	*Group V (aromatic)*
Basmati 6936_IRRI	F63	Pakistan	*Group V (aromatic)*
Basmati 122_IRRI	F76	Pakistan	*Group V (aromatic)*
Blomberg	2264	India	*Group V (aromatic)*
Chinisaldkar	2258	India	*Group V (aromatic)*
Tararori Basmati	2278	India	*Group V (aromatic)*
Pusa basmati-1	2279	India	*Group V (aromatic)*
HBC-18	2280	India	*Group V (aromatic)*
HARAND 379	3174	Pakistan	*Group V (aromatic)*
Basmati 802	3232	Pakistan	*Group V (aromatic)*
BASPUNJAB			*Group V (aromatic)*
SADRI	3175	Iran	*Group V (aromatic)*
BOKEHMWE	3186	Myanmar	*indica*
DIGAN			*tropical japonica*
YUN AROMATIC GLUTINOUS	3222	China	*indica*
RTS12	136	Vietnam	*indica*
Nang Thom Chc Dao	3127	Vietnam	*indica*
Chao do	3147	Laos	*indica*
HNN	3155	Laos	*indica*
GUOR LAO	3167	Vietnam	*indica*
PAWSANHMWE	3191	Myanmar	*indica*
Champa Tong 54_IRRI	F21	Thailand	*indica*
Ratnaweo	132	Sri Lanka	*indica*
MAWAINE CHN	3169	Myanmar	*indica*

Cell color indicates polymorphisms that either match (yellow) or do not match (blue) the NERICA1 haplotype.
All polymorphisms above 5% frequency in this germplasm set were concatenated to construct these haplotypes.
Only markers with informative polymorphisms (above 5% frequency) are shown in this table.
The proposed region of introgression from WAB63-81 into NERICA1 is outlined in a red box

REFERENCES

Ahn S.N., Bollich C.N., Tanksley S.D. (1992) RFLP tagging of a gene for aroma in rice. Theoretical and Applied Genetics 84:825-828

Bhattacharjee P., Singhal R.S., Kulkarni P.R. (2002) Basmati rice: a review. International Journal of Food Science & Technology 37:1-12

Bradbury L., Henry R., Jin Q., Reinke R.F., Waters D.L.E. (2005a) A perfect marker for fragrance genotyping in rice. Molecular Breeding 16:279-283

Bradbury L.M.T., Fitzgerald T.L., Henry R.J., Jin Q., Waters D.L.E. (2005b) The gene for fragrance in rice. Plant Biotechnology Journal 3:363-370

Buttery R.G., Ling L.C., Juliano B.O. (1982) 2-acetyl-1-pyrroline: An important aroma component of cooked rice. Chem Ind (London) 12:958-959

Chen S., Wu J., Yang Y., Shi W., Xu M. (2006) The *fgr* gene responsible for rice fragrance was restricted within 69 kb. Plant Science 171:505-514

Chen S., Yang Y., Shi W., Ji Q., He F., Zhang Z., Cheng Z., Liu X., Xu M. (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. Plant Cell 20:1850-1861

Fitzgerald M.A., Sackville Hamilton N.R., Calingacion M.N., Verhoeven H.A., Butardo V.M. (2008) Is there a second fragrance gene in rice? Plant Biotechnology Journal 6:416-423

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638

Ikeda R., Sokei Y., Akintayo I. (2007) Reliable multiplication of seed for NERICA varieties of rice, *Oryza sativa* L. Genetic Resources and Crop Evolution 54:1637-1644

Itani T. (1993) History, cultivation, and breeding of aromatic rice cultivars. Bull Hiroshima Pref Univ 5:267-281

Jodon N.E. (1944) The inheritance of flower fragrance and other characters in rice. Journal of the American Society of Agronomy 36:844-848

Jones M., Dingkuhn M., Aluko/snm G., Semon M. (1997) Interspecific *Oryza Sativa* L. X *O. Glaberrima* Steud. progenies in upland rice improvement. Euphytica 94:237-246

Kadam B.S., Patankar V.K. (1938) Inheritance of aroma in rice. Chronica Botanica 4:32

Kovach M.J., Calingacion M.N., Fitzgerald M.A., McCouch S.R. (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). Proceedings of the National Academy of Sciences of the United States of America 106:14444-14449

Lorieux M., Petrov M., Huang N., Guiderdoni E., Ghesquière A. (1996) Aroma in rice: genetic analysis of a quantitative trait. Theoretical and Applied Genetics 93:1145-1151

Qiu Z.J., Zhang Y.S. (2003) Why fragrant rice produced in Thailand can be sold worldwide? World Agric (China) 2:33-36

Semagn K., Ndjiondjop M.N., Lorieux M., Cissoko M., Jones M., McCouch S.R. (2007) Molecular profiling of an interspecific rice population derived from a cross between WAB56-104 (*Oryza sativa*) and CG14 (*Oryza glaberrima*). African Journal of Biotechnology 6:2014-2022

Shi W., Yang Y., Chen S., Xu M. (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. Molecular Breeding 22:185-192

Sweeney M., McCouch S. (2007) The complex history of the domestication of rice. Annals of Botany 100:951-957

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D., McCouch S.R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3:e133

Takano-Kai N., Jiang H., Kubo T., Sweeney M., Matsumoto T., Kanamori H., Padhukasahasram B., Bustamante C., Yoshimura A., Doi K., McCouch S. (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics 182:1323-1334

Vaughan D.A., Miyazaki S., Miyashita K. (2004) The rice gene pool and human migrations. In: Werner D. (ed) Biological Resources and Migration. Springer, Berlin, pp 1-13

Yamanaka S., Nakamura I., Watanabe K., Sato Y.-I. (2004) Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theoretical and Applied Genetics 108:1200-1204

CHAPTER 6:
MAPPING THE GENETIC DETERMINANT OF FRAGRANCE IN KAI NOI LEUANG

INTRODUCTION

A rice breeder in Laos, Chanthakhone Boualaphanh, undertook a breeding project to introgress fragrance from a highly fragrant Laotian landrace, Kai Noi Leuang (KNL), into an elite non-fragrant Laotian variety, Thasano1 (TSN1). Through collaboration with the International Rice Research Institute (IRRI) in the Philippines, it was discovered that KNL did not possess the *badh2.1* allele (Fitzgerald et al., 2008), known to be the predominant genetic determinant of fragrance in rice (Chen et al., 2008). Further investigations found that KNL possessed a much higher level of 2AP than most normal fragrant varieties, and that it appeared to retain 2AP content during storage, unlike varieties whose 2AP content decreases significantly over time (data not shown). The *BADH2* gene was later sequenced in KNL to determine if it possessed one of the other *BADH2* alleles predicted to cause fragrance (Kovach et al., 2009). No coding mutation was found in KNL that could explain the presence of fragrance, although two different sequencing primer sets consistently failed to amplify in KNL (BadhapG3 and BadhapG4). It was hypothesized that KNL may possess a lesion in a different gene in the metabolic pathway leading to the production of 2AP (Kovach et al., 2009). Given the importance of the fragrance trait for modern rice breeders (especially in Southeast Asia), discovering a novel fragrance gene would be a unique finding that might allow for refinement of the 2AP phenotype in rice varieties. This study was therefore designed to map the genetic determinant of fragrance in KNL while simultaneously exploring the usefulness of a new medium-resolution, high-throughput SNP genotyping platform for rice.

MATERIALS AND METHODS

Plant materials

In her efforts to introgress fragrance from KNL to TSN1, Chanthakhone created a BC_4F_2 population, with TSN1 as the recurrent parent, selecting in each generation for fragrant individuals to backcross. The result was a set of ~190 BC_4F_2 plants, which were grown in the field at IRRI in 2009. DNA was collected from these backcross individuals, and gas chromatography was used to measure 2AP content from leaf tissue from each plant (2AP content of mature grains was not measured). If a 2AP peak appeared on the gas chromatogram, the accession was designated "fragrant", while in the absence of a 2AP peak, the accession was designated as "non-fragrant".

Genotyping

KNL, TSN1, and 68 BC_4F_2 plants (33 fragrant; 35 non-fragrant) were initially genotyped with a 384–plex Illumina BeadXpress SNP genotyping platform developed to maximize the number of polymorphic SNPs between the *Indica* and *Japonica* varietal groups (S. McCouch & K. Zhao, Cornell University, personal communication). Since KNL is a *tropical japonica* and TSN1 is an *indica* variety, this assay was predicted to provide a large number of polymorphic markers that could be used for preliminary mapping of fragrance in this segregating population. Raw intensity data from the BeadXpress assay were analyzed using the ALCHEMY allele calling algorithm (M. Wright, Cornell University, personal communication). DNA for the parents and 68 BC_4F_2 lines was also sent to Cornell University, where it was used to confirm the Illumina BeadXpress data (in collaboration with M. Fitzgerald and D. Daygon from IRRI).

RESULTS AND DISCUSSION

Mapping fragrance in KNL

In the BC_4F_2 population, fragrance segregated as 3 non-fragrant : 1 fragrant, suggesting fragrance was inherited as a single gene in this population. Out of the 384 SNPs in the BeadXpress assay, 219 were found to be polymorphic between KNL and TSN1 and could be used for mapping. Despite evidence of a single gene, a standard QTL analysis method was used to establish any significant marker/phenotype associations. The phenotype was coded as "1" for fragrant accessions and "0" for non-fragrant accessions and was treated as a nominal trait. The marker data were input into the QTL statistical software QGene, and a single significant QTL was detected on chromosome 8 near SNP id8005226. Visual inspection confirmed that in many of the lines designated as "fragrant", there was a KNL introgression at this region of chromosome 8 (Figure 6.1). Since SNP id8005226 is located at 19.7 Mb and the *BADH2* gene is located at 20.3 Mb on chromosome 8, these mapping results strongly suggested that the genetic determinant of fragrance in KNL was due to a defective *BADH2* gene and not a novel genetic factor. To confirm the BeadXpress results, three of the previously described sequencing markers near *BADH2* (Kovach et al., 2009) were used to screen the 68 BC_4F_2 plants (one marker in the *BADH2* gene and two flanking markers). This sequencing data confirmed the validity of the BeadXpress allele calls in all cases (data not shown), demonstrating the accuracy of this genotyping platform/allele calling algorithm. It was noted that despite the strong association between this region of chromosome 8 and fragrance in this population, nearly two-thirds of the individuals designated as "fragrant" either lacked any KNL introgression at *BADH2* (13/34) or were heterozygous for the introgression (8/34), both of which would be predicted to be non-fragrant. In other words, only 12 of the

34 “fragrant” accessions were homozygous for the KNL introgression at *BADH2* and would be predicted to be fragrant.

Analysis of *BADH2* in KNL

To date, the *BADH2* gene is the only genetic factor that has been shown to be responsible for the production of 2AP in rice. This fact, coupled with the mapping data, suggested that the *BADH2* gene in KNL was defective, and that previous sequencing efforts had failed to find the causal mutation. Therefore, primers were designed to sequence across the *BADH2* gene, with particular focus on the region of *BADH2* that failed to amplify in the previous study (Kovach et al., 2009).

The sequencing results initially failed to find any new coding mutations in KNL relative to Nipponbare (functional *Badh2*). Yet, a region spanning Exon 4 and part of Exon 5 consistently failed to amplify in KNL. To explain why this region failed to amplify, it was hypothesized that a large genomic perturbation must exist in that region. Therefore, PCR was attempted using primers outside the problem region (Ex4_F / Ex5_R), resulting in successful amplification in KNL, although the PCR product size was dramatically smaller than expected (~500 bp). When amplified in TSN1 (actually, no TSN1 DNA was sent to Cornell, so a BC_4F_2 individual that was homozygous TSN1 at *BADH2* was used), the PCR product was of the expected size (~1300 bp) (Figure 6.2). These results were consistent with there being a large deletion in this region of the *BADH2* gene in KNL. The amplicons were then sequenced to determine the exact nature of the disturbance in KNL, which identified an 806 bp deletion in KNL relative to Nipponbare. This deletion encompasses part of Exon 4 and Exon 5, and results in a premature STOP codon and truncation of the

Figure 6.1: Illumina BeadXpress Results for TSN1 x KNL BC_4F_2

This graphic displays the SNP allele calls from the BeadXpress assay run on TSN1, KNL, and 33 fragrant/35 non-fragrant BC_4F_2 plants for the 15 polymorphic SNPs on chromosome 8. The plant identity is listed on the left, followed by the fragrance phenotype, followed by the allele calls for each SNP. A SNP allele is colored red if the genotype matched the TSN1 parent, blue if the genotype matched the KNL parent, or purple if heterozygous (missing data are colored white). The relative position of the *BADH2* gene is depicted with a vertical yellow line. Note the apparent KNL introgression near the *BADH2* gene in many of the "fragrant" individuals.

		id8000131	id8000315	id8000470	id8000666	id8001426	id8001667	id8002103	id8002632	id8004692	id8005226		id8006032	id8006321	id8006792	id8006885	id8007878
		8	8	8	8	8	8	8	8	8	8		8	8	8	8	8
TSN1		AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
KNL		BB	AA	AA	BB	BB	BB	--	AA	--	AA		BB	BB	AA	AA	AA
606	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AA	AA	BB
639	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AA	AA	BB
610	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AA	AB	BB
631	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AA	AB	BB
637	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AB	AB	BB
609	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		BB	BB	AB	AB	BB
622	Frag	AB	AB	AB	AB	AA	AA	BB	BB	BB	AA		AA	AA	BB	BB	BB
625	Frag	BB	AA	AA	BB	AA	AA	BB	BB	BB	AA		AA	AA	BB	BB	BB
626	Frag	AB	AB	AB	AB	AA	AA	BB	BB	BB	AA		AA	AA	BB	BB	BB
629	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		AA	AA	BB	BB	BB
668	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AA		AA	AA	BB	BB	BB
643	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		BB	BB	AA	AA	BB
633	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AB	AB	AA	AA	BB
634	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AB	AB	AA	AA	BB
636	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AB	AB	AB	AB	BB
640	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AB	AB	AB	BB	BB
501	Frag	AA	BB	BB	AB	AB	AB	AB	--	BB	AB		AB	AB	BB	BB	BB
638	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AB	AB	BB	BB	BB
504	Frag	AA	BB	BB	AA	AB	AB	AB	AB	BB	AB		AB	AB	BB	BB	BB
635	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AA	AA	BB	BB	BB
664	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	--		AA	AA	BB	BB	BB
505	Frag	AA	BB	BB	AB	AA	AA	BB	BB	--	BB		AA	AA	BB	BB	BB
529	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
539	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
557	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
558	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
560	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
577	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
588	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
619	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
632	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
665	Frag	AA	BB	BB	AA	AA	AA	BB	BB	AB	BB		AA	AA	BB	BB	BB
666	Frag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
670	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	AB		AA	AA	BB	BB	BB
530	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
523	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
444	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
447	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
448	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
461	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
464	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
506	Nonfrag	AA	BB	BB	AB	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
512	Nonfrag	AA	BB	BB	AB	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
531	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
540	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
545	Nonfrag	AB	AB	AB	AB	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
546	Nonfrag	BB	AB	AB	AB	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
548	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
553	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
554	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
562	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
563	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
564	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
571	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
587	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
590	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
592	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
593	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
603	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
611	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
616	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
617	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
649	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
652	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
658	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
679	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
680	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB
684	Nonfrag	AA	BB	BB	AA	AA	AA	BB	BB	BB	BB		AA	AA	BB	BB	BB

BADH2 protein, eliminating the majority of the aldehyde dehydrogenase domain (Figure 6.3). Both the BadhapG3_R and BadhapG4_F primers from the previous study were anchored in this deletion, which explains why those amplicons invariably failed in KNL.

This study identified a new allele of the *BADH2* gene from KNL, *badh2.11*, which has an 806 bp deletion that severely disrupts the coding region of the gene and results in a putatively nonfunctional *BADH2* protein. One additional fragrant variety from Vietnam, Khau Tan Luong (KTL), was found to have an identical *badh2.11* allele.

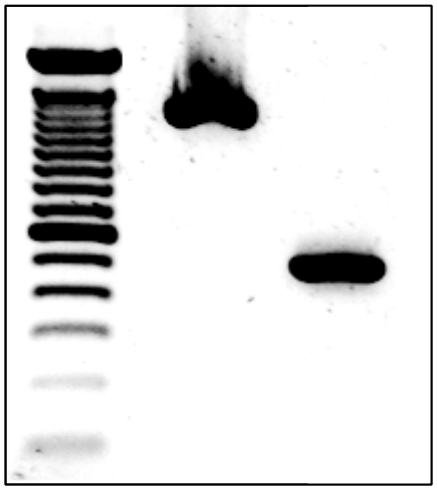


Figure 6.2: Ex4_F / Ex5_R Amplification in KNL. When primers Ex4_F and Ex5_R were amplified in KNL (fragrant) and TSN1 (non-fragrant), the product sizes were vastly different. TSN1 had the expected product size of 1296 bp (based on the Nipponbare sequence), while the product in KNL was only ~500 bp, indicating a large deletion in the *BADH2* gene in KNL. Lane 1 = Kb+ DNA Ladder; Lane 2 = TSN1; Lane 3 = KNL

Figure 6.3: Genomic DNA and predicted protein of *Badh2* (wild-type) and *badh2.11* (fragrant) alleles

(A) Primers Ex4_F (pink) and Ex5_R (light blue) were used to amplify and sequence a part of the *BADH2* coding region in KNL. The sequence data identified an 806 bp deletion in KNL relative to Nipponbare, encompassing parts of Exon 4 (yellow) and Exon 5 (green).
(B) This deletion causes the predicted BADH2 protein in KNL to be truncated from 503 amino acids (wild-type) to 135 amino acids, eliminating the majority of the aldehyde dehydrogenase domain

A)

Nipponbare (*Badh2* allele; Non-Fragrant):

```
GATTGTGGGAAGCCTCTTGATGAAGCAGCATGGGACATGGTATGTGGCCAGTTATCCACTGTATGAATATGTAGTTGCCTACACAGCAATCTTTCCTGAACATGAATCCTG
ATGTATGATATTCCATTTGTCAGGACGATGTTGCTGGATGCTTTGAGTACTTTGCAGATCTTGCAGAATCCTTGGACAAAAGGCAAAATGCACCTGTCTCTCTTCCAAT
GGAAAACTTTAAATGCTATCTTCGGAAAGAGCCTATCGGTGTAGTTGGGTTGATCACACCTTGGTATTTCACATTTTTCTCTCATCCTGCGCTTATATTTATTTATGAC
CCAAGCATGGTACTAAATAGTACTAGTAACATGCATATACTGAATGAGTTTACAACTTTACATGATTTTTTTGAACTATGAAAGTTGAAGACATTTGAGATTTTATTCC
TCTTCTCTTGTGCAAACATATTATTGTCTCACAAATTGTACCTAGCAGCTACTCTCTCCGTTTCATATTATAAGTCGTTTGACTTTTTTCCTAGTCAAAATGTGTTAAG
TTTGACCAAGTTTATAGAAAAATTTAGCAACATCTAAAATATCAAAGTCATGTTTTAGTGTTTTTTCAGGCTCTCATGTAAGCAATTTTGATGTGCCCTCTCCTTTCTT
CTTAATATAATGATACACAGCTCTTGTGTATTCAAAGGAAAATATATATATATATAATGATACACACCTCTCCTCCGTGTTAATGCAGCTCATTTGTTCTGTCCCGGTT
CAAATATCTATTTTTCTCATATGTTGTCAGCATGATTCACTTAATTTAGTATATAGAAGATGCCATTATTTATGTCTGGAATCTTACTGCAGAAGGGAAAACAATTGAT
AACGGAATTGATTGCATTCTAATTTGTTGTTTCTTTGTTATGTTCTTATCGACAATTACAAATTTGATTCTGAGAATCATGTTCGGGATGTGTATTTCTACTGCAGGAA
CTATCCTCTCCTGATGGCAACATGGAAGGTAGCTCCTGCCCTGGCTGCTGGCTGTACAGCTGTACTAAAACCATCTGAATTGGCTTCCGTGTAAGTTTAACATGTTAAC
TTGTTAATGTCATACCCATGCTAGTTGCAATGACATTTGATTTTAAAATGTTGTGGCATGTCCATGCTGCAAGCAATGTAATTTGAAATCTCTCTCTATCATTAATTAC
CAGGACTTGTTTGGAGCTTGCTGATGTGTGTAAAGAGGTTGGTCTTCCTTCAGGTGTGCTAAACATAGTGACTGGATTAGGTTCTGAAGCCGGTG
```

KNL (*badh2.11* allele; Fragrant):

```
GATTGTGGGAAGCCTCTTGATGAAGCAGCATGGGACATGGTATGTGGCCAGTTATCCACTGTATGAATATGTAGTTGCCTACACAGCAATCTTTCCTGAACATGAATCCTG
ATGTATGATATTCCATTTGTCAGGACGATGTTGCTGGATGCTTTGAGTACTTTGCAGATCTTGCAGAATCCTTGGACAAAAG..........................
..........................................................................................................
.....GCAACATGGAAGGTAGCTCCTGCCCTGGCTGCTGGCTGTACAGCTGTACTAAAACCATCTGAATTGGCTTCCGTGTAAGTTTAACATGTTAAC
TTGTTAATGTCATACCCATGCTAGTTGCAATGACATTTGATTTTAAAATGTTGTGGCATGTCCATGCTGCAAGCAATGTAATTTGAAATCTCTCTCTATCATTAATTAC
CAGGACTTGTTTGGAGCTTGCTGATGTGTGTAAAGAGGTTGGTCTTCCTTCAGGTGTGCTAAACATAGTGACTGGATTAGGTTCTGAAGCCGGTG
```

B)

Nipponbare BADH2 Protein (*Badh2* allele; Non-Fragrant)

```
MATAIPQRQLFVAGEWRAPALGRRLPVVNPATESPIGEIPAGTAEDVDAAVAAAREALKRNRGRDWARAPGAVRAKYLRAIAAKIIERKSELARLETLDCGKPLDEAAW
DMDDVAGCFEYFADLAESLDKRQNAPVSLPMENFKCYLRKEPIGVVGLITPWNYPLLMATWKVAPALAAGCTAVLKPSELASVTCLELADVCKEVGLPSGVLNIVTGLG
SEAGAPLSSHPGVDKVAFTGSYETGKKIMASAAPMVKPVSLELGGKSPIVVFDDVDVEKAVEWTLFGCFWTNGQICSATSRLILHKKIAKEFQERMVAWAKNIKVSDPL
EEGCRLGPVVSEGQYEKIKQFVSTAKSQGATILTGGVRPKHLEKGFYIEPTIITDVDTSMQIWREEVFGPVLCVKEFSTEEEAIELANDTHYGLAGAVLSGDRERCQRL
TEEIDAGIIWVNCSQPCFCQAPWGGNKRSGFGRELGEGGIDNYLSVKQVTEYASDEPWGWYKSPSKL-
```

KNL BADH2 Protein (*badh2.11* allele; Fragrant)

```
MATAIPQRQLFVAGEWRAPALGRRLPVVNPATESPIGEIPAGTAEDVDAAVAAAREALKRNRGRDWARAPGAVRAKYLRAIAAKIIERKSELARLETLDCGKPLDEAAW
DMDDVAGCFEYFADLAESLDKRQHGR-
```

An allele-specific marker for *badh2.11*

This new *BADH2* allele in KNL may be targeted for selection and introgression into elite rice varieties in the future. To facilitate this task, an allele-specific marker was designed to amplify the *badh2.11* allele and allow resolution of the 806bp deletion easily on an agarose gel. The marker creates a 192 bp amplicon in individuals carrying a homozygous *badh2.11* allele, a 1000 bp amplicon in individuals having a homozygous wild-type allele at *BADH2*, and heterozygotes will exhibit both bands (Figure 6.4).

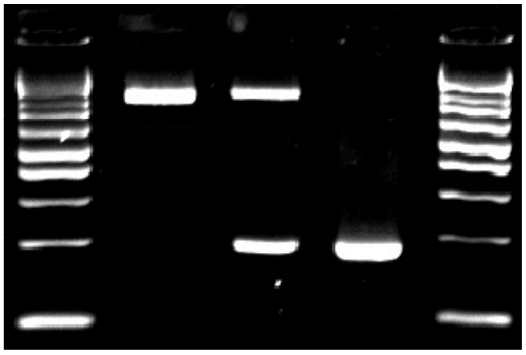


Figure 6.4: *badh2.11* Allele-Specific Marker. A marker for the *badh2.11* allele was designed to amplify a large band (1000 bp) in individuals homozygous for the wild-type *Badh2* allele and a small band (194 bp) in individuals homozygous for the *badh2.11* allele. Lanes 1 & 5 = Kb+ DNA Ladder; Lane 2 = TSN1; Lane 3 = Heterozygote; Lane 4 = KNL

Unresolved Issues

While this study successfully identified a new fragrance allele of the *BADH2* gene, the mapping discrepancies remain unresolved. If the genetic determinant of fragrance in KNL is due to the *BADH2* gene, then why would 21 of the 34 accessions from the mapping study that were predicted to be non-fragrant (due to the lack of a

homozygous KNL introgression) have positive 2AP peaks in the gas chromatogram? If only the heterozygotes showed small 2AP peaks, it might indicate a dosage effect at this gene, yet this was not the case. Another possible explanation would be that there are additional genes contributing to the lack of 2AP accumulation in this population. This also seems unlikely, since the original BC_4F_2 population segregated very closely to the expected 3:1 ratio for a single gene. The final possible explanation for this discrepancy could be that the phenotyping results are incorrect, possibly due to a mislabeling error or contamination during gas chromatography analysis. The results of past studies suggest that all rice plants produce very low but detectable levels of 2AP, with non-fragrant plants producing less than the threshold level for the human nose (Buttery et al., 1983; Chen et al., 2008). It has also been shown that the aerial parts of the rice plant produce more 2AP than the grain (Yoshihashi et al., 1999). Since leaves were used for measuring the 2AP phenotypes in the BC_4F_2 population, it may be that this low level of 2AP production is detectable in the leaves, but not in the mature grains of "non-fragrant" plants. Further analysis of this BC_4F_2 population will attempt to reconcile the genotypic and phenotypic data. Additional studies will also attempt to map modifiers of 2AP accumulation in a population that is fixed for the *badh2.11* allele, although this will require extremely sensitive and reproducible measurements of 2AP levels in the grain.

REFERENCES

Buttery R.G., Ling L.C., Juliano B.O., Turnbaugh J.G. (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. Journal of Agricultural and Food Chemistry 31:823-826

Chen S., Yang Y., Shi W., Ji Q., He F., Zhang Z., Cheng Z., Liu X., Xu M. (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. Plant Cell 20:1850-1861

Fitzgerald M.A., Sackville Hamilton N.R., Calingacion M.N., Verhoeven H.A., Butardo V.M. (2008) Is there a second fragrance gene in rice? Plant Biotechnology Journal 6:416-423

Kovach M.J., Calingacion M.N., Fitzgerald M.A., McCouch S.R. (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). Proceedings of the National Academy of Sciences of the United States of America 106:14444-14449

Yoshihashi T., Huong N.T.T., Kabaki N. (1999) Quality evalutation of Khao Dawk Mali 105, an aromatic rice variety of northeast Thailand. JIRCAS Working Report 30:151-160

CHAPTER 7:

The Genetic Control of Black Hull in *Oryza sativa*

Introduction

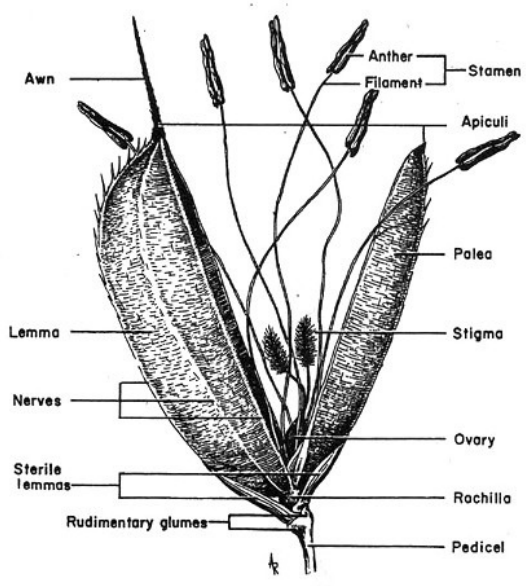
Background

The genetic mechanisms underlying the suite of traits altered during rice domestication (the "domestication syndrome") have been extensively studied, and many of the genes and causal mutations responsible for these traits have been identified and characterized. While many of the domestication syndrome traits exhibit quantitative inheritance, evidence to date suggests they are often subject to at least one major genetic determinant, with a series of modifying elements that further attenuate the trait (Gepts, 2004; Doebley et al., 2006). Therefore, major genes have been found that explain a large proportion of the variation for morphological traits such as plant architecture (Jin et al., 2008; Tan et al., 2008), grain size (Fan et al., 2006; Shomura et al., 2008; Wang et al., 2008; Takano-Kai et al., 2009), grain shattering (Konishi et al., 2006; Li et al., 2006; Lin et al., 2007), and flowering time (Kojima et al., 2002; Tadege et al., 2003; Doi et al., 2004).

In addition to these traits, which tended to make the plant more agronomically acceptable or more nourishing to humans, the rice domestication process also led to the elimination of pigmentation from most plant organs. The inheritance of the pigmentation patterns in various rice organs has been characterized (Parnell et al., 1921; Chao, 1928; Takahashi, 1957), and in some cases the underlying genes have been identified (Sakamoto et al., 2001; Saitoh et al., 2004; Sweeney et al., 2006; Furukawa et al., 2007). While the evolutionary significance of these color abolitions is not always clear, the consistent loss of pigmentation across cultivated crop species

suggests that this was an important factor during the domestication of major crop species.

In rice, the grain is composed of an exterior hull or "husk" (sometimes referred to as the glume), which covers the caryopsis or "seed". The hull is comprised of sterile lemmas, rachilla, palea, and lemma, with the lemma covering two-thirds of the seed and the edges of the palea fitting into the lemma, providing a tight sheath around the caryopsis (Figure 7.1). The hull tissue is maternally derived, and so its phenotype is influenced solely by the genotype of the mother plant. During panicle and grain development, the hulls of Asian wild rice (*Oryza rufipogon*) exhibit a progression from green to black coloration, with the black color appearing at physiological maturity. In contrast, the hulls of nearly all cultivated rice varieties (*Oryza sativa*) fail to accumulate the black compounds upon maturity, but remain a light tan/straw to sometimes golden or even brown color (for the purposes of this study, any non-black hull phenotype will be referred to as "light").



Chang & Bardenas, 1965)

Figure 7.1: Detailed Anatomy of a Rice Spikelet; Pre-Fertilization

Classical Genetics of Black Hull Inheritance

The inheritance of hull color in rice has been examined on multiple occasions dating back to the early part of the last century (Parnell et al., 1921). A variety of hull pigmentation patterns have been identified, including straw, yellow, furrowed, piebald, golden, red, purple, brown, blackish green, blackish brown, and black. These designations often seem somewhat subject in the literature, and little consensus exists between studies. Yet, it is clear from classical genetic studies that brown, red, and purple hull coloration are inherited independently of black hull coloration. Thus the focus of this report will be solely on black hull coloration, since this is the distinguishing feature between the hulls of *O. sativa* and its wild progenitor. Parnell (1921) reported both 3:1 and 9:7 ratios for the inheritance of black hull in different crosses, implicating a single dominant gene or the interaction between two complementary genes for this phenotype. Other classical inheritance studies reported similar results, pointing to the interaction of two epistatic genes (Chao, 1928; Mitra and Ganguli, 1936; Kuang et al., 1946; Kuriyama and Kudo, 1967; Tripathi and Rao, 1979; Shobha Rani et al., 2008), or the possible involvement of a third gene when crosses are made between the *Indica* and *Japonica* varietal groups (Nagao and Takahashi, 1954; Rao and Seetharaman, 1973; Maekawa, 1982, 1984). It was also demonstrated that one of the black hull factors was linked with a factor controlling the production of awns (Kuang et al., 1946; Maekawa, 1982), an interesting feature that will be returned to later. Also of note is that two studies on the inheritance of black hull coloration in *O. barthii* reported a 9:7 segregation ratio (Nwokeocha, 1998; Aladejana, 2000), offering the possibility that the same genes affect this trait across multiple A-genome *Oryza* species.

The *Ph* gene: Genetic Determinant of Phenol Reaction

In the middle part of the century, researchers in Japan discovered that when certain varieties of *O. sativa* with straw-colored (light) hulls were soaked in a phenol solution, a chemical reaction occurred that turned the hulls black (Kondo and Kasahara, 1940; Kasahara, 1941). It was shown that this phenol reaction could be used to discriminate between rice varieties, particularly between *Japonica* and *Indica*. When soaked in phenol solution, *Indica* varieties tended to turn black, while *Japonica* varieties remained unchanged (Oka, 1953). The inheritance of this trait was later found to be controlled by a single dominant gene (Moringa et al., 1943; Nagao and Takahashi, 1952). This led to the hypothesis that the phenol reaction gene (*Ph*) was involved in the natural process whereby hulls turn black at maturity, accompanied by the action of a second gene (*Bh*; *black hull*), which produced the phenol substance upon which *Ph* acted (Kuriyama and Kudo, 1967). The action of *Ph* was thought to be an oxidation of the phenolic compounds, altering them from colorless to the characteristic black color found in mature hulls. Maekawa (1982) found that varieties with black hulls invariably exhibited a positive phenol reaction, substantiating a possible link between black hull coloration and the phenol reaction.

The *Ph* gene was later mapped to the long arm of rice chromosome 4 (McCouch et al., 1988; Saito et al., 1991; Lin et al., 1994) and the biochemical role of the *Ph* gene in phenol oxidation was confirmed upon the recent cloning and characterization of *Phr1* (Yu et al., 2008). *Phr1* encodes a polyphenol oxidase (PPO), a class of enzymes known to be involved in discoloration of plant organs upon ripening or cellular damage. Therefore, in addition to being responsible for the phenol reaction used to distinguish *Indica* and *Japonica* varieties, the researchers implicate *Phr1* as a cause of the slow grain discoloration that tends to occur during the storage of some rice

varieties. Transformation of a phenol reaction-negative accession with the functional *Phr1* gene resulted in the restoration of a positive phenol reaction in the hull (Yu et al., 2008). While there are three *Phr1*-like homologs in the rice genome, none of these homologs were predicted to be functional, and *Phr1* was shown to be the only source of PPO activity in rice. Three different coding mutations were found in *Phr1*: an 18 bp deletion, a 29 bp deletion, and a 1 bp insertion, all resulting in nonfunctional alleles and a negative phenol reaction. All three of these mutations were detected in *Japonica* varieties, while nearly all *Indica* and wild accessions had a functional *Phr1*. Evolutionary analysis suggested that the three defective *Phr1* alleles arose independently in the recent past in *Japonica*, where at least one of the alleles (18 bp deletion) appears to have been under positive selection and has reached high frequency in *Japonica*. The authors postulate that a lack of PPO activity may be preferred in *Japonica* where discoloration during storage is culturally distasteful, while potential disease resistance mechanisms imparted by oxidized phenolic compounds may have encouraged the retention of PPO activity in *Indica*, which inhabits warmer climates (Yu et al., 2008). This would help to explain the clear divergence between *Indica* and *Japonica* at *Phr1*, although the true reason for the dichotomy at this locus remains unclear.

The *Bh* Genes

As stated previously, most classical inheritance studies predicted either two or three complementary genes controlling the black hull phenotype. When three genes were implicated, they were designated *Bh-a*, *Bh-b*, and *Bh-c*, with *Bh-c* corresponding to the phenol reaction gene, *Ph* (now *Phr1*) (Maekawa, 1984). Early investigations of the *Bh-a* and *Bh-b* genes focused mainly on the geographical distribution of varieties possessing these defective alleles. The *Bh-a* allele was found at a high frequency

(79.6%) and appeared to be widespread in varieties from India, through Southeast Asia and China, into Japan, while the *Bh-b* allele was found at low frequency and was localized within a small region of Japan near Hokkaido (Maekawa, 1982, 1984). This suggested that while the *Bh-a* allele was widespread in contemporary *O. sativa* landraces, the *Bh-b* allele remained isolated. Inheritance relationships between the three genes indicated that *Bh-a* and *Bh-b* assorted independently from each other, as did *Bh-b* and *Ph*. Yet, *Bh-a* and *Ph* significantly deviated from random assortment, indicating that these two genes may be linked (Maekawa, 1984).

The chromosomal locations of the *Bh-a* and *Bh-b* genes have not yet been confirmed. One paradoxical study using wild trisomic lines suggested that there was in fact a single *Bh* gene and it was located on chromosome 12 (Jena and Khush, 1990), although no other study has mapped the black hull trait to this location. Another study using a mapping population derived from a cross between Pei-kuh (*O. sativa*; *indica*) and wild accession W1944 (*O. rufipogon*) identified a single gene segregating for black hull coloration on chromosome 5, which they speculated was *Bh-b* (Cai et al., 2002). This location for *Bh-b* was corroborated in a different cross between Nipponbare (*O. sativa*; *temperate japonica*) and Kasalath (*O. sativa*; *aus*) (Harushima et al., 1998). No study to date has reported a genetic location of the *Bh-a* gene.

Two more recent studies investigated the inheritance of weedy characteristics in *O. sativa* (weedy meaning "wild-like"), including hull coloration. In crosses between a breeding line derived from CO39 (*O. sativa*; *indica*) and several weedy strains, black hull color was found to have a very high heritability (90%) (Gu et al., 2005c). A total of three QTLs were detected, on chromosomes 1, 4, and 7, with the QTL on chromosome 4 (*qHC4*) explaining over 50% of the variation for black hull (Gu et al.,

2005a; Gu et al., 2005c). The QTL on chromosomes 1 and 7 explained only minor proportions of the phenotypic variation in hull color, and were therefore referred to as modifiers of the major QTL on chromosome 4. Since the *Phr1* gene (*Bh-c*) exists within the marker interval for this QTL on chromosome 4, it was assumed that this QTL was allelic to *Bh-c* (Gu et al., 2005c).

Goals of This Study

Black hull color was arguably one of the first morphological traits to be investigated by classical rice geneticists. Yet, nearly a century after the dawn of genetic theory, a complete genetic explanation for the inheritance of black hull in rice remains incomplete. The Yu et al. (2008) study confirmed that *Phr1* (formerly *Ph*) controls the change in hull coloration upon exogenous treatment with a phenol solution, but did not make any link between *Phr1* and the natural occurrence of black hull coloration *in planta*. In fact, while classical genetics studies speculated that *Ph* played a role in black hull coloration, no direct evidence supporting this hypothesis has been presented (Maekawa, 1982). Also, the genetic location and identity of the *Bh-a* gene has yet to be reported. The dual purpose of the current study was therefore to map the *Bh-a* gene and also to determine if the *Phr1* gene is necessary for the black hull phenotype.

Materials and Methods

Plant Materials

Three separate mapping populations were used simultaneously in this study, each contributing unique information, toward the goal of rapidly mapping the location of the factors controlling black hull coloration in rice.

1) BS125 (*O. sativa*; *indica*) x *O. longistaminata* accession WL02; BC_1 population (henceforth referred to as the SL population). This interspecific population was created nearly two decades ago, and was used to construct one of the first saturated genetic maps in rice (Causse et al., 1994). As *O. longistaminata* is the most distantly related member of the A-genome *Oryza* species (Ren et al., 2003), substantial sterility barriers forced nearly 10,000 crossing attempts before a single viable F_1 seed was obtained. This F_1 plant was subsequently backcrossed to BS125, resulting in the 113 BC_1 lines that were used to create the genetic map. These backcross plants have been maintained vegetatively in the Cornell greenhouses for the last 17 years. During that period, if a BC_1 plant produced viable seed (BC_1F_2 seed) the seed was collected and stored. Many of the plants never produced viable BC_1F_2 seed, likely a manifestation of the sterility barriers present in this wide interspecific cross. Also, over the course of time, many of the BC_1 plants either died or were recalcitrant to vegetative progragation and were lost. At the time when this population was investigated for mapping black hull coloration, only 60 BC_1 plants remained in the greenhouse. Fifty-seven out of the original 113 BC_1 plants produced viable seed that could be phenotyped for hull coloration (a set that only partially overlapped with the remaining live BC_1 plants). Since BS125 possessed light hulls while WL02 possessed black hulls, and since substantial genotypic data were already available for this population due to the genetic map construction, this population was attractive as a starting point for mapping the rough location of the genetic cause of black hull.
2) Cybonnet (*O. sativa*; *tropical japonica*; NSF-TV #397) x *O. rufipogon* accession 506A (NSF-TV #506A) (henceforth referred to as the Cyb x 506A population). This population was originally created as part of a CSSL library construction project. Based on currently available genotypic data, the *O. rufipogon* parent

clusters with the *aus* subpopulation. Cybonnet possesses light hulls while 506A has black hulls, and the F_2 seed had black hulls. This population entered the black hull mapping project as a set of remnant F_2 seed, of which 205 individuals were planted and grown in the Cornell greenhouses in spring/summer 2009. Later, as informative recombinants were identified, F_3 families of those individuals were planted for fine mapping.

3) Cybonnet (*O. sativa*; *tropical japonica*; NSF-TV #397) x *O. rufipogon* accession 549A (NSF-TV #549A) (henceforth referred to as the Cyb x 549A population). This population was also originally created as part of a CSSL library construction project. The currently available genotypic data cluster 549A with an *O. rufipogon*-specific group being internally referred to as "Independent 1". The 549A parent possesses black hulls and the F_2 seed had black hulls. This population entered the black hull mapping project as both a set of remnant F_2 seed and also a set of BC_1 seed that had been backcrossed to the Cybonnet parent. One hundred sixty-six BC_1 individuals were planted and grown in the Cornell greenhouses in spring/summer 2009. Later, as informative recombinants were identified, BC_1F_2 families of those individuals were planted for fine mapping.

In addition, a set of *O. sativa* and *O. rufipogon* accessions were selected for association analysis based on their hull color phenotype. These accessions were selected from the purified germplasm set that was created under NSF-TV Award #0606461.

Genotyping and Data Analysis

The SL population had already been genotyped at 726 markers (of all types; mostly RFLPs (Causse et al., 1994)), providing more than enough genotypic data for an initial

mapping of the black hull trait. These marker data were contained on old floppy disks, and no computer that could read the disks could be found at Cornell. Luckily, a collaborator from Korea was able to send the genotypic data for the SL population, and after some de-convolution, the genotypic data were able to be matched to the seed from the BC_1 plants that had been collected at Cornell. Mapping could only be performed using the accessions for which both genotypic and phenotypic (hull coloration) data were available, leaving 52 BC_1 individuals out of the original population of 113. Despite the qualitative nature of this trait, basic QTL mapping procedures were used to identify statistically significant genotype/phenotype associations. The hull coloration phenotype was coded as "1" for black hull or "0" for light hull. A rudimentary QTL mapping script was used to analyze the SL population data in the R statistical software package. After this initial round of mapping, an additional 4 MITE markers (based on MITE transposon polymorphisms (G. Wilson, Cornell University, personal communication) and 4 SSR markers were added to the target region to confirm the RFLP data.

Initially, for both the Cyb x 506A and Cyb x 549A populations, SSR markers in the target region were chosen and surveyed for polymorphism between the parents using either 3% agarose or 4% polyacrylamide gel electrophoresis. Markers that were polymorphic between all three parents were then used to screen both the 205 F_2 plants (Cyb x 506A) and 166 BC_1 plants (Cyb x 549A). After informative recombinants were identified, additional markers were created to further delineate the recombination breakpoints. These markers included either SSRs or newly-designed indel markers. To create new codominant indel markers, the published Nipponbare sequence (*Japonica*) was aligned to the shotgun sequence contigs for 93-11 (*Indica*), and indels of 20 or more base pairs were targeted for marker development. Primers were

designed to amplify regions of ~100-200 base pairs surrounding the indel (to allow resolution of the polymorphisms on 3% agarose gels). The assumption for this indel marker creation method was that both *O. rufipogon* parents used in this study were more closely related to *Indica* than *Japonica*, and so indels detected between Nipponbare and 93-11 had a high chance of being detected in both crosses with Cybonnet (a *tropical japonica*). Approximately half of the designed indel markers were found to amplify and be polymorphic between Cybonnet and the *O. rufipogon* parents, suggesting this indel marker design method could be moderately efficient in similar crosses.

Fine Mapping

From the original set of 205 F_2 and 166 BC_1 plants, those individuals found to have recombination breakpoints in the target region were planted in either F_3 or BC_1F_2 families for further fine mapping. Additionally, some individuals that were heterozygous across the entire target region were also selected for screening in the next generation. A total of 907 F_3 individuals and 604 BC_1F_2 individuals were screened for informative recombination breakpoints; 1511 plants were screened in total. As new recombinants decreased the size of the target region containing *Bh-a*, additional markers were created flanking the narrower target. When no polymorphic SSRs or indels were available, a SNP-based genotyping assay was used. First, ~700 base pair regions were amplified and sequenced in the parents. When a SNP that distinguished Cybonnet from both wild parents was detected, the SNP was genotyped in potential recombinants using the KBiosciences Competitive Allele Specific PCR SNP genotyping system (KASPar). FRET intensity values were obtained using the Allelic Discrimination function on an ABI 7900H machine at the Cornell Life

Sciences Core Laboratories. Clustering and analysis of KASPar allele calls was performed using the SDS2.1 software.

Results

Initial Mapping in the SL Population

Initial mapping of black hull coloration was performed using the SL population. Out of the 57 BC_1 plants that produced viable BC_1F_2 seed over the past two decades, 35 had dark hulls and 22 had light hulls. Since many of the BC_1 plants were lost over the course of time or never produced viable seed, segregation ratios from this population are likely to be inaccurate. Nevertheless, the observed phenotypic ratio is statistically consistent with a 1:1 ratio ($p = 0.085$), which would indicate the action of a single gene for black hull in this population.

Out of the 57 BC_1 individuals that produced seed, there were only 52 individuals for which the genotypic information could be confidently assigned, and therefore 52 individuals were used for QTL analysis. The results of the QTL analysis with the R software indicated a single genomic region associated with hull coloration on chromosome 4, with the QTL peak reaching infinity (a complete association between one of the RFLP markers and the qualitative phenotype). Visual inspection of the marker data confirmed a block of linked markers on chromosome 4 that were highly associated with hull coloration, with RFLP marker RZ740 (at ~24.7 Mb) perfectly associated with the phenotype. The addition of 4 polymorphic MITE markers and 4 SSRs flanking RZ740 confirmed the RFLP data. Additionally, recombination breakpoints were detected between RG788 and RZ740 in one BC_1 line, and between RZ740 and RM6823 in another, suggesting the target gene must be downstream of marker RG788 (~20.1 Mb) and upstream of marker RM6823 (~26.2 Mb) (Figure 7.2).

Thus, using phenotypic data from seeds produced by the BC_1 plants and genotypic data from the Causse (1994) study, the location of a major genetic determinant of black hull color in this population could be mapped to an approximately 6 Mb region of chromosome 4.

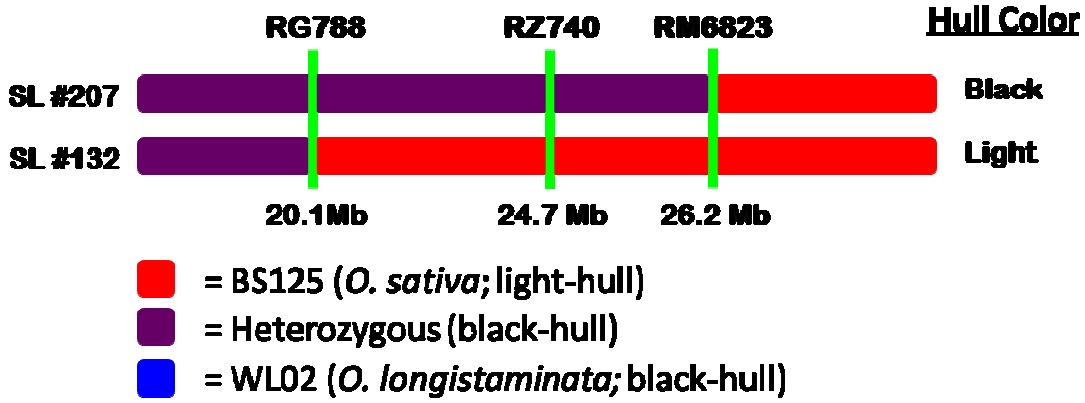


Figure 7.2: Informative Recombinants for Hull Color from SL Population

Incidentally, this region of chromosome 4 did not contain the *Phr1* gene (31.6 Mb), which has been suggested as one of the genetic determinants of black hull coloration in rice. This result makes sense, though, since the *O. sativa* parent of the SL population was an *indica* variety. Almost all *Indica* varieties were previously shown to possess a functional *Phr1* gene, and so we would not have expected this locus to be segregating in this cross.

Mapping in the Cybonnet x *O. rufipogon* Populations

Two interspecific populations between *O. rufipogon* and a common *O. sativa* parent (cv. Cybonnet; *tropical japonica*) were used to further dissect the location of the genetic cause of black hull coloration on chromosome 4. First, several polymorphic SSR markers flanking the ~6 Mb target region identified from the SL population were screened in Cybonnet, the two wild parents, and both the F_2 and BC_1F_2 populations.

The results of this initial survey indicated a strong association between the presence of an *O. rufipogon* allele in this region and the incidence of black hull in the mapping populations. This suggested that the genetic determinant of black hull coloration on chromosome 4 was the same in the SL, Cyb x 506A, and Cyb x 549A populations.

The genotypes of several individuals in the target region immediately indicated an anomaly. While almost all individuals with the *O. rufipogon* allele across the target region had black hulls, there were some individuals that possessed the *O. rufipogon* segment, but had light hulls. This presented a paradox, and indicated the possibility that a second gene controlling black hull coloration was segregating in the Cyb x 506A and Cyb x 549A populations. Further analysis showed that individuals that recombined to the Cybonnet allele near the 3' end of the target region almost always had light hull color. Fortuitously, it was actually at this point that the recent publication describing the *Phr1* gene was discovered (Yu et al., 2008). Given that the *Phr1* gene is almost always nonfunctional in *Japonica* varieties, it would be logical that Cybonnet possessed one of the three frameshift-inducing mutations in *Phr1*. Therefore, markers were designed to amplify the two prevalent deletions in *Phr1* described previously (Yu et al., 2008). These markers were screened in Cybonnet and the two wild parents, but Cybonnet was not found to possess either deletion. To test if Cybonnet had a dysfunctional *Phr1* gene, grains of Cybonnet were soaked in a 2% phenol solution for 3 days, along with several other *Japonica* and *Indica* controls. The Cybonnet seeds exhibited a negative phenol reaction, indicating that Cybonnet does indeed have a dysfunctional *Phr1*. So sequencing primers were designed and the *Phr1* gene was sequenced in Cybonnet and the wild parents. This revealed that Cybonnet possesses a 1 bp insertion in the 1st exon of *Phr1*, which is the third of the previously described mutations in *Phr1*, and had only been found in 1 variety (Yu et al., 2008). A

polymorphic SSR near the *Phr1* gene (RM5771) was then screened in the two mapping populations to determine whether *Phr1* was required for the black hull phenotype. The results indicated that both a functional *Phr1* and a second factor contained within the 6 Mb *O. rufipogon* segment were required for the black hull phenotype. Any recombinants that possessed the *O. rufipogon* segment at the target region, but recombined to Cybonnet at *Phr1* had light hulls. Similarly, recombinants that were Cybonnet at the target region, but had a functional *Phr1* (the *O. rufipogon* allele) also had light hulls. The only individuals with black hulls had an *O. rufipogon* introgression at both the target region and *Phr1*. This result proves for the first time that *Phr1*, classically referred to as *Bh-c* or *Ph*, is indeed required for the black hull phenotype in addition to its requirement for a positive phenol reaction. It also strongly suggested that *Bh-a*, which had not yet been mapped, was within the 6 Mb target region upstream of *Phr1* on chromosome 4.

Having now determined the requirement of a functional *Phr1* for black hull coloration, all subsequent recombinants were first screened with the SSR tightly linked to *Phr1*, and only those recombinants with the *O. rufipogon* allele at *Phr1* were retained. With the second factor fixed, *Bh-a* could now be mapped as a single gene. Several hundred F_3 or BC_1F_2 individuals were planted and screened at a time, and informative recombinants were retained. Additional markers (SSR, indel, KASPar) were designed to further delineate the recombination breakpoints and narrow down the target region containing *Bh-a*. The target region was eventually narrowed down to an approximately 90 kb region between indel marker FM22.748 and RM3558 (Figure 7.3).

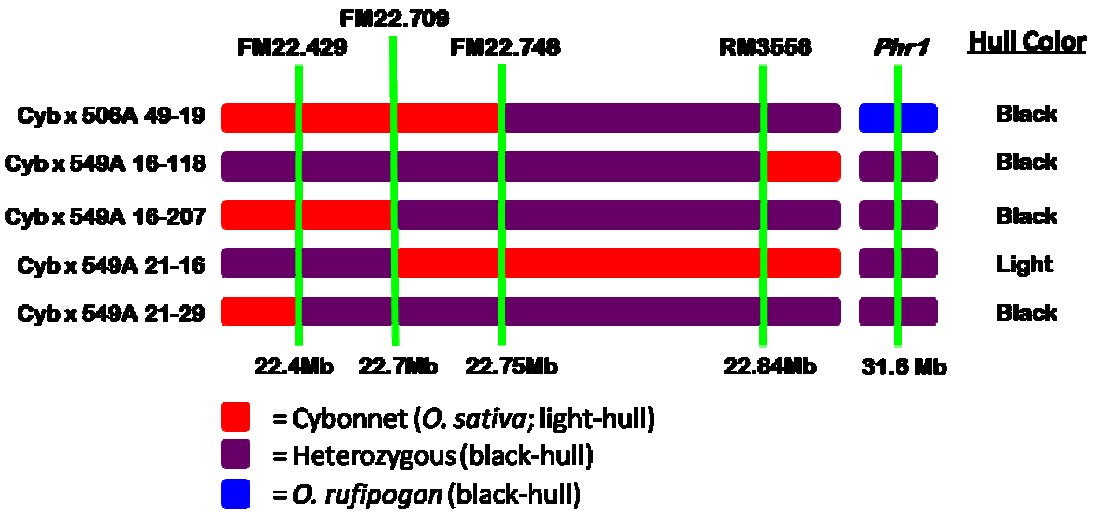


Figure 7.3: Informative Recombinants for Hull Color from Cyb x 506A and Cyb x 549A Populations

This 90 kb region contained 14 predicted open reading frames, including three with no annotated function (Table 7.1). Among the 14 ORFs, there were two amino acid transporters and a third predicted protein with homology to the amino acid transporter family, all three located in tandem (with the predicted protein in the middle). Since amino acid transporters are known to be integral membrane-bound proteins involved in the transport of a variety of compounds across cellular membranes, these appeared to be strong candidates for *Bh-a*. Contrary to prevailing hypotheses suggesting *Bh-a* encoded a protein involved in the biosynthesis of phenolic compounds (Kuriyama and Kudo, 1967), the current hypothesis was that *Bh-a* transports phenolic compounds to the appropriate cellular compartment in the hull tissue, where they contribute to the black hull phenotype.

Table 7.1: 14 ORFs within Fine-Mapped *Bh-a* Target Region

Gene/Marker	Position (MSU6)	Annotated Function
FM22.748	22,748,226	INDEL Marker
LOC_Os04g38600	22,751,948	glyceraldehyde-3-phosphate dehydrogenase
LOC_Os04g38610	22,754,032	RAN guanine nucleotide release factor
LOC_Os04g38620	22,758,069	NAP domain containing protein
LOC_Os04g38630	22,762,697	helicase conserved C-terminal domain containing protein
LOC_Os04g38640	22,769,737	OsDegp5 - Putative Deg protease homologue
LOC_Os04g38650	22,777,064	expressed protein
LOC_Os04g38660	22,784,737	transmembrane amino acid transporter protein
LOC_Os04g38670	22,786,688	amino acid transporter family protein
LOC_Os04g38680	22,792,706	transmembrane amino acid transporter protein
LOC_Os04g38690	22,797,272	membrane associated DUF588 domain containing protein
LOC_Os04g38700	22,798,633	molybdopterin biosynthesis protein CNX3
LOC_Os04g38710	22,806,341	hypothetical protein
LOC_Os04g38720	22,809,444	no apical meristem protein
LOC_Os04g38730	22,816,387	hypothetical protein
RM3558	22,838,859	SSR Marker

To identify potential functional polymorphisms in these three genes, the coding regions were sequenced in Cybonnet and the two wild parents. Several polymorphisms were detected that distinguished the Cybonnet allele from the two *O. rufipogon* alleles. The most interesting of these was a 22 bp deletion in the predicted amino acid transporter gene (LOC_Os04g38670), which was in between the other two

amino acid transporter genes. This predicted protein was subjected to an inter-species BLAST search, which identified putative homologs in maize, sorghum, grape, soybean, and castor bean. The protein with the highest similarity was an unmapped maize protein (LOC_100285277), which was annotated as "amino acid/polyamine transporter II" and had been found very recently in a large-scale cDNA sequencing study (Alexandrov et al., 2009). The rice protein had very high homology with this maize protein, except that the rice protein only covered the C-terminal end of the maize protein. Upon further analysis, the amino acid transporter just upstream of LOC_Os0438670 (LOC_Os0438660) had extremely high homology to the N-terminal portion of the maize protein, suggesting that these two ORFs actually encoded a single protein product. It was discovered that by replacing the 22bp deletion that was missing in LOC_Os04g38670 in Cybonnet, the gene prediction software now predicted that LOC_Os0438660 and LOC_Os04g38670 were a single gene, encoding a protein with 77% homology to maize protein LOC_100285277. This finding mandates a re-annotation of these genes; the actual wild-type gene is 2465 bp from start to stop codon, has three exons and one intron (of 1166 bp), with a predicted cDNA of 1191 bp and protein of 396 amino acids (Figure 7.4). The wild-type protein is predicted to contain 10 transmembrane domains, suggesting this is an integral membrane-bound protein. The 22bp deletion in the 3rd exon of this gene in Cybonnet is predicted to truncate the amino acid transporter protein prior to the last 4 transmembrane domains, putatively rendering it nonfunctional (Figure 7.5).

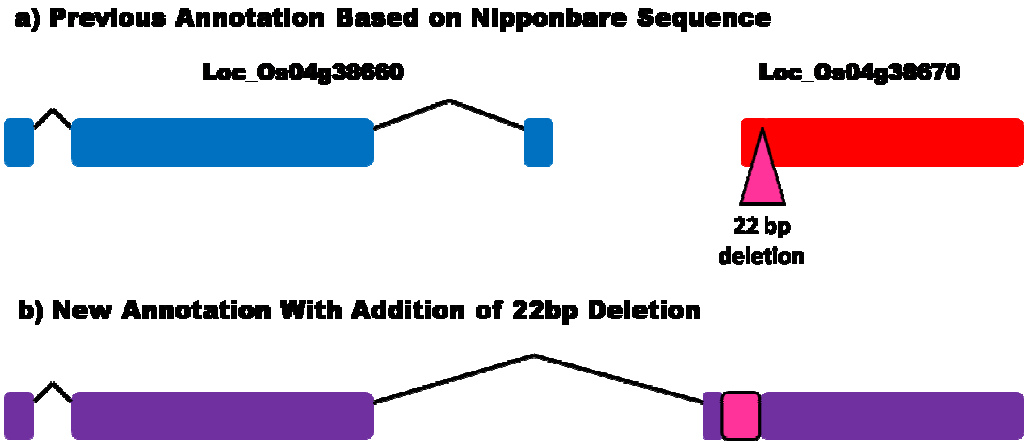


Figure 7.4: Structure of Loc_Os04g38660 and Loc_Os04g38670 With and Without 22bp Deletion

a) The MSU6 build of the Nipponbare genome detected two adjacent genes on chromosome 4, Loc_Os04g38660 and Loc_Os04g38670, both with homology to amino acid transporters. Sequencing Loc_Os04g38670 in Cybonnet, 506A and 549A identified a 22 bp deletion in Cybonnet (and Nipponbare) relative to the wild accessions.

b) When the 22 bp deletion is accounted for, FGENESH gene prediction software then identifies a single contiguous gene with high homology to a maize amino acid transporter (Loc_100285277).

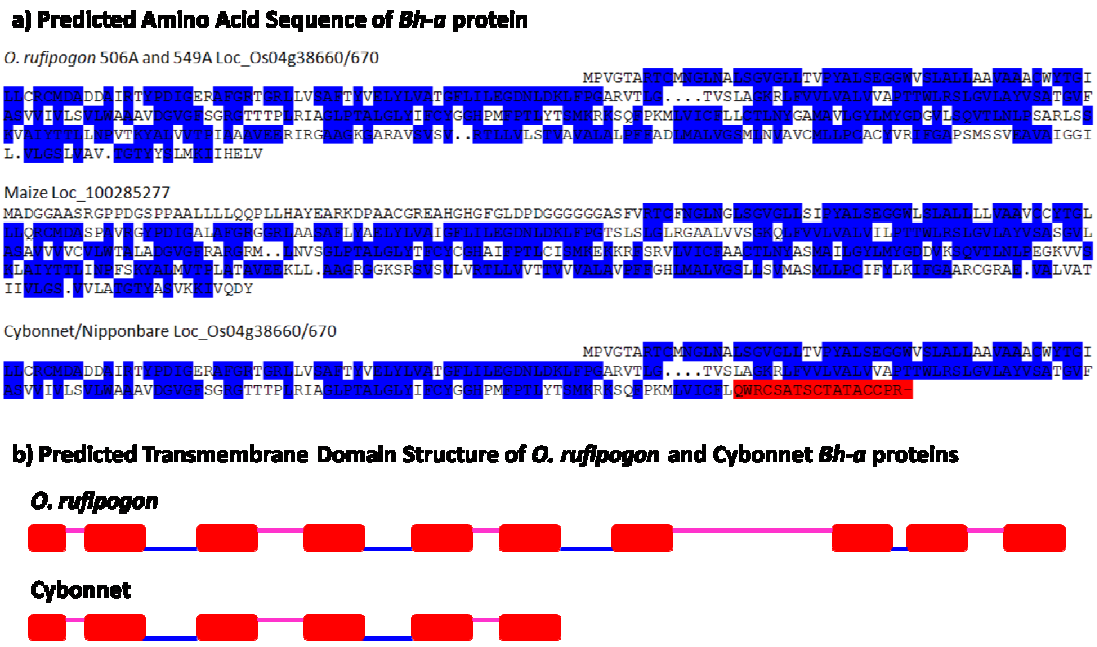


Figure 7.5: *Bh-a* Protein Comparisons

a) The amino acid sequences of the newly annotated amino acid transporter protein are compared between *O. rufipogon*, Cybonnet/Nipponbare, and the putative maize ortholog. Residues that are identical between the rice and maize proteins are highlighted in blue. The 22 bp deletion in Cybonnet/Nipponbare causes a frameshift and premature truncation of the protein, which is highlighted in red.

b) Using the "Transmembrane helices in proteins v. 2.0" program, the structure of the predicted *O. rufipogon* and Cybonnet proteins are compared. Red boxes represent putative transmembrane domains and blue and pink lines represent domains predicted to exist inside or outside, respectively.

Development of an allele-specific marker for the 22 bp deletion

To facilitate high-throughput genotyping of the 22-bp deletion in this amino acid transporter gene, an allele-specific marker was developed. This marker amplifies a 190 bp fragment in wild-type (black hull) individuals, and a 168 bp fragment in individuals with the 22 bp deletion (light hull), allowing easy allelic discrimination by running the PCR products on a 3% agarose gel (Figure 7.6).

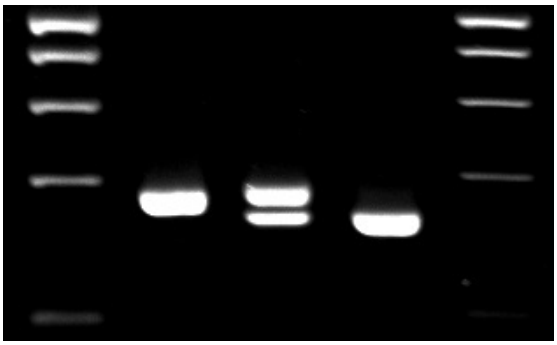


Figure 7.6: Allele-specific marker for 22 bp deletion. An allele-specific marker was created to allow rapid genotyping of the *Bh-a* gene on an agarose gel. Lane 1 & 5 = Kb+ DNA Ladder; Lane 2 = *O. rufipogon* 506A; Lane 3 = Heterozygote; Lane 4 = Cybonnet

Confirmation Via Association Analysis

In silico sequence analysis determined that Nipponbare, 93-11, and Cybonnet all possessed the 22 bp deletion in LOC_Os04g38670 relative to *O. rufipogon*. To determine whether this 22 bp deletion is diagnostic of light hull color, an association analysis was performed. A large panel of diverse germplasm was assembled, including 435 *O. sativa* landraces/varieties and 148 *O. rufipogon*/*O. nivara*/*O. spontanea* accessions. Of the *O. sativa* accessions, 431 had light hulls and 4 had black hulls. The inverse situation was true of the wild accessions, where 128 had black hulls and 20 had light hulls (Supplemental Table 7.1). These 583 accessions were assayed using the

allele-specific marker for the 22 bp deletion. For the wild accessions, all of those with black hulls (N = 128) had the wild-type (non-deletion) allele, while all of the light-hulled accessions (N = 20) possessed the 22 bp deletion. Similarly, all of the black-hulled *O. sativa* accessions tested (N = 4) had the wild-type (non-deletion) allele. Of the 431 *O. sativa* accessions with light hulls, 399 possessed the 22 bp deletion (93%), but there were 32 accessions with light hulls that did not have the deletion. This presented the possibility that the 22 bp deletion was not the functional polymorphism responsible for the light hull phenotype. Alternatively, since a functional *Phr1* gene was shown to be required for black hull, this discrepancy could be due to a loss of function at the *Phr1* gene in these 32 accessions. To test this hypothesis, the *Phr1* gene was partially sequenced in these 32 accessions to assay for the presence of the three previously reported functional mutations in this gene (Yu et al, 2008). The sequencing results identified an 18 bp deletion in the *Phr1* gene in 25 out of the 32 accessions. Interestingly, 24 out of these 25 light-hulled accessions lacking the 22 bp deletion in *Bh-a* but having a nonfunctional *Phr1* gene were from the *temperate japonica* subpopulation (with one accession from the *aus* subpopulation). Overall, these results demonstrated that 98% of all light-hulled accessions of *O. sativa* tested in this study could be explained by either the 22 bp deletion in LOC_Os04g38670 or an 18 bp deletion in the *Phr1* gene. The remaining 2% (N = 8) of light-hulled accessions that possessed neither the 22 bp deletion nor any of the previously reported functional polymorphisms in *Phr1* were all members of the *aus* subpopulation. The current hypothesis is that these eight light-hulled *aus* accessions possess a unique polymorphism in either LOC_Os04g38670 or the *Phr1* gene, disrupting the pathway leading to black pigment accumulation in the hull. This situation would be remarkably similar to the evolutionary history of the *Rc* gene controlling pericarp color in rice, where a unique nonfunctional allele was found at a very low frequency within the *aus*

subpopulation (Sweeney et al., 2007). Further analysis at the sequence and expression levels of both genes will be necessary to confirm this hypothesis. Alternatively, these accessions could possess a novel genetic lesion in the pathway leading to the black hull phenotype, and would therefore provide the starting point for mapping studies to identify this novel factor.

The results of this association analysis demonstrated that the 22 bp deletion in this amino acid/polyamine transporter on rice chromosome 4 was predictive of the light-hull phenotype in 93% of the *O. sativa* and 100% of the *O. rufipogon* accessions tested. Additionally, in almost all cases, the light hull phenotype in *O. sativa* could be predicted by the presence of the 22 bp deletion in LOC_Os04g38670 or a nonfunctional *Phr1* gene, with the exception of eight accessions from the *aus* subpopulation. The conclusion is therefore that an amino acid/polyamine transporter gene (a concatenation of LOC_Os04g38660 and LOC_Os04g38670) on rice chromosome 4 is *Bh-a* and the wild-type form is required for the black hull phenotype in rice. A 22 bp deletion in the 3rd exon of this gene is predicted to render the protein nonfunctional, resulting in the light hull phenotype that is typical of *O. sativa*.

Discussion

The genetic basis of black hull in rice

Many of the genes underlying traits involved in the rice domestication syndrome have been identified, allowing us to form a more coherent history of how rice evolved into one of the world's most important staple crops. While some of the traits that were altered by human selection have an obvious connection to improving the rice plant agronomically, the reason black hull coloration was selected against is less clear. Nevertheless, the fact that nearly all wild rice species possess black hulls, while nearly

all *O. sativa* landraces and modern varieties lack black hulls, indicates a close relationship between this trait and the process of rice domestication. The purpose of this study was to dissect the inheritance of black hull color in rice and to map the genes responsible for this phenotype. By learning more about this trait, which is associated with the rice domestication syndrome, we hoped to add one more piece to the puzzle of rice domestication and improve our understanding of the genetic mechanisms responsible for domestication traits in crop species.

Classical genetic studies identified three complementary genes controlling the inheritance of black hull color in rice, *Bh-a*, *Bh-b*, and *Bh-c*. The *Bh-c*, or *Ph*, gene was the first to be mapped (McCouch et al., 1988) and cloned (Yu et al., 2008) due to its association with an important varietal classification method, positive phenol reaction. While *Bh-c* was thought to be necessary for the black hull phenotype, this had never been demonstrated conclusively until the current study. *Bh-b* appears to be a relatively rare allele (Maekawa, 1982), and its limited distribution in Japan suggests it is an allele restricted to the *temperate japonica* subpopulation, although more work would be necessary to confirm this. It is also possible that *Bh-a* and *Bh-b* are allelic, although the mapping studies that report *Bh-b* on chromosome 5 would refute this (Harushima et al., 1998; Cai et al., 2002). The location and identity of *Bh-a* was elucidated in the current study, revealing that *Bh-a* and *Bh-c* are both located on the long arm of rice chromosome 4. The linkage between these two genes may have inhibited efforts to genetically map *Bh-a*, since the two genes are in coupling phase linkage and therefore would be inherited together more often than expected. On the other hand, almost all inheritance studies reported a 9:7 segregation ratio in the F_2 generation of various black hull x light hull crosses, suggesting that while these two genes are linked, the linkage is weak enough that they segregate almost independently.

The average physical distance per centiMorgan in the rice genome is ~244 kb/cM (Chen et al., 2002). The *Bh-a* and *Phr1* genes are ~8750 kb apart on chromosome 4, which would correspond to a genetic distance of ~36 cM. A rough estimation of recombination frequency between *Bh-a* and *Phr1* based on empirical data from this study (actual number of recombinants/total number screened) was 43.1%, which would indicate the actual genetic-to-physical distance in this region is larger than the genome-wide average.

It is interesting to note that early inheritance studies reported a linkage relationship between a factor for black hull coloration and the production of awns, a long hair-like extension of the epidermis at the tip of the lemma (Figure 7.1). The recombination frequency between the two factors was reported as 20.8 (Kuang et al., 1946) and 28.9 (Maekawa, 1982). A separate fine mapping project in the McCouch lab has narrowed the location of the awn factor on chromosome 4 to ~13.4 Mb. With the *Bh-a* gene located at 22.8 Mb, it is likely that these inheritance studies were detecting linkage between *Bh-a* and the awn factor. *Phr1* is located at 31.6 Mb, and so would be less likely to show linkage with the awn factor. This example is presented to illustrate the curious phenomenon that rice domestication genes appear to be clustered together in the genome (Xiong et al., 1999; Cai et al., 2002; Ji et al., 2006). In addition to those controlling black hull coloration and awns, there are several other domestication genes located on the long arm of chromosome 4. The *Sh4* gene encodes a transcription factor responsible for abscission layer formation at the junction of the seed and the panicle. The functional form of the gene in *O. rufipogon* causes the seed to shatter and fall from the panicle, while a SNP in the gene in *O. sativa* dramatically reduces the incidence of seed shattering (Li et al., 2006). *Sh4* is located at 34 Mb on chromosome 4, and given the large linkage blocks in an F_2/F_3 or BC_1F_2 population,

this would explain why nearly all black-hulled individuals from the mapping populations used in this study exhibited severe seed shattering. The *GIF1* gene encodes a cell wall invertase that acts as a key regulator of starch synthesis in the endosperm during grain filling, likely contributing to rice yield increases during domestication (Wang et al., 2008). *GIF1* is located at 20.2 Mb on chromosome 4, closely linked to the *Bh-a* gene. In this case, the linkage relationship is in repulsion, so the domesticated form would have required recombination between these tightly linked genes to obtain both light hulls and improved grain filling. A gene controlling the production of rhizomes, a trait responsible for the transition of wild rice from a perennial to an annual habit, was mapped to between RM119 and RM273 (21.2 – 22.5 Mb) on chromosome 4 (Hu et al., 2003), which would be very near the *Bh-a* gene. A variety of QTL controlling domestication and yield-related traits in rice have also been mapped to this region of chromosome 4. In summary, the region between 20 and 34 Mb on rice chromosome 4 appears to harbor a large number of genes associated with the rice domestication syndrome, including the major shattering and both major black hull genes.

Models for the biochemical basis of black hull

The identity of *Bh-a* as an amino acid/polyamine transporter does not agree with the previous hypothesis postulated by several researchers that *Bh-a* would encode a gene involved in the biosynthesis of phenolic compounds in the rice hull (Kuriyama and Kudo, 1967; Rao and Seetharaman, 1973). This hypothesis was probably based on the apparent complementary action between *Bh-a* and *Bh-c*, and the knowledge that *Bh-c* controlled the oxidation of phenol in rice hulls upon exogenous treatment with phenol solution. If *Bh-c* oxidizes phenols applied to the exterior, turning them black, it would make sense that *Bh-a* would be involved in the production of those phenols within the

rice hull. Yet, a new model for the biochemical basis of black hull coloration must now be developed, since *Bh-a* does not encode an enzyme directly involved with the biosynthesis of phenolic compounds. Two alternative models are proposed:

1) *Phr1* is predicted to have an N-terminal signal peptide that would target it to the secretory pathway (Yu et al., 2008). This is in contrast to the majority of PPO literature, which states that PPO enzymes are normally targeted to the chloroplast (see (Mayer, 2006) for review). Using the subcellular prediction tool TargetP, *Bh-a* is not predicted to be targeted to the secretory pathway, chloroplast, or mitochondria, although the signal peptide prediction tool SignalP does predict a signal peptide targeting *Bh-a* to the secretory pathway. Thus, no conclusions can be made at this time regarding the subcellular location of the *Bh-a* protein, and the location of the *Phr1* protein is also ambiguous. Nevertheless, one model for the molecular basis of black hull is that *Bh-a* is involved with the transport of polyphenolic compounds into the appropriate cellular compartment, where they are then oxidized by the action of the polyphenol oxidase (PPO) encoded by *Phr1*.
2) As a membrane-bound transport protein, *Bh-a* would not likely be involved directly in the biosynthesis of polyphenolics. Yet, it is possible that the *Bh-a* protein transports one of the precursors of the polyphenolic compounds into the appropriate cellular compartment, therefore indirectly affecting the biosynthesis of polyphenolics.

Given that *Indica* varieties turn black upon treatment with phenol solution, either of the above models support the idea that polyphenol synthesis is ubiquitous in both *O. rufipogon* and *O. sativa*, but the compounds (or precursors) fail to be transported to appropriate cellular compartment for oxidation by PPO (*Indica*) or both fail to be transported and are not oxidized due to lack PPO activity (*Japonica*). Additional

studies will be necessary to elucidate the chemical composition of the polyphenolic compounds that accumulate in black-hulled varieties, as well as investigations into the biochemical interactions between the polyphenols and the *Bh-a* and *Phr1* proteins.

Evolutionary significance

The domestication-related genes that have been cloned and characterized in rice thus far can be divided into two classes: "pan-*O. sativa*", where the derived allele has reached near-fixation in the cultivar, or "population-specific", where the derived allele remained isolated within one varietal group or subpopulation. In the case of black hull coloration, the *Bh-a* gene appears to have followed the first evolutionary path, while the *Phr1* gene followed the later.

Since the 22 bp deletion in *Bh-a* was detected in nearly all light-hulled varieties of *O. sativa*, this mutation was presumably selected early during the domestication of *O. sativa* or has been under recent, strong positive selection in both varietal groups. Whether the deletion occurred in *O. rufipogon* and was then selected from standing variation in the wild, or whether the deletion arose after the initial domestication of *O. sativa* remains to be investigated. Either way, the near-fixation of the dysfunctional allele in *O. sativa* suggests the light-hull trait has been strongly preferred by humans. The fact that *O. barthii* (Nwokeocha, 1998; Aladejana, 2000), *O. longistaminata* (this study), and most *O. rufipogon*/*O. nivara* accessions (this study) have an intact *Bh-a* gene suggests the black-hull phenotype is advantageous under natural conditions. Those *O. rufipogon* accessions with light hulls are likely the result of introgression events from the cultivar at this region of chromosome 4. Confirmation of this hypothesis could be achieved by performing haplotype analysis across the *Bh-a* genomic region. Haplotype analysis would also allow discernment of the origin of the

22 bp allele to determine whether the deletion arose in a common ancestor to both varietal groups or if it arose in one varietal group and was transferred via introgression into the other, consistent with the evolutionary histories of several other domestication genes in rice (Kovach et al., 2007; McCouch et al., 2010).

While the 22 bp deletion in *Bh-a* is common across *O. sativa*, the *Phr1* gene has undergone divergent selection in the *Indica* and *Japonica* varietal groups, as evidenced by the high frequency of functional *Phr1* in *Indica* and nonfunctional *Phr1* in *Japonica*. If the *Bh-a* mutation arose early during rice domestication, and this defect alone was sufficient to abolish the accumulation of polyphenolic compounds in the rice hull, why or how was the *Phr1* gene selected upon? This question will require further genetic and biochemical analysis to adequately answer, but it presents an intriguing evolutionary puzzle. It suggests that the PPO activity imparted by *Phr1* is required under the conditions *Indica* varieties are traditionally grown, while it could be lost in *Japonica* varieties without evolutionary cost. This means that a functional *Phr1* must impart some unknown pleiotropic effect in addition to its contribution to black hull coloration.

The ultimate question regarding the evolutionary significance of this trait is: Why would humans have selected against black rice hulls? Polyphenolic compound accumulation has been shown to offer a wide array of disease and pest resistances to plants (Li and Steffens, 2002; Mayer, 2006; Melo et al., 2006), and so the loss of these resistances must have been worth the agronomic gain of having light hulls. One potential explanation for what appears to have been strong selection against black hull coloration in the emerging cultivar would be the marked connection between seed pigmentation and seed dormancy across the cereal crop species. Seed dormancy is an

essential trait for most wild plant species, since it prevents germination until environmental conditions are favorable, decreases competition between individuals of the same species, and allows a species to survive natural catastrophes (Finkelstein et al., 2008). Loss of dormancy and seed shattering have been referred to as the two most important features of the domestication syndrome in seed-propagated crop species (Gepts, 2004).

There are essentially two types of seed dormancy in plants, physiological and physical (Finch-Savage and Leubner-Metzger, 2006), which in rice are referred to as embryo (kernel)-imposed or seed coat-imposed. Seed coat-imposed dormancy in rice is due to the inhibition of germination by the hull and the testa (pericarp) layer of the caryopsis. This inhibition may be either due to a physical barrier provided by the covering tissues or by inhibitor chemicals found within these tissues (Bewley and Black, 1982). Studies have long shown that pigmentation of the seed coat has an inhibitory effect on seed germination, meaning unpigmented seeds possess considerably lower levels of dormancy (particularly in wheat) (Gfeller and Svejda, 1960; Khan and Strand, 1977; DePauw and McCaig, 1983; Debeaujon et al., 2000; Groos et al., 2002; Gu et al., 2006; Debeaujon et al., 2007).

In three different crosses between light-hulled, non-dormant and black-hulled, dormant rice accessions, hull color and dormancy were significantly correlated, ranging from 25 to 59% (Gu et al., 2005b). Other studies have reported a similarly strong correlation between these two traits in rice (Gu et al., 2005a; Gu et al., 2005c). As further evidence that the black hull trait and dormancy are interrelated, several dormancy QTL mapping studies have identified a QTL for dormancy located at the very same region of chromosome 4 that contains *Bh-a* and *Phr1*. The amount of

variance in dormancy explained by this QTL on chromosome 4 was relatively low, ranging from 6% (Gu et al., 2005c) to 11% (Gu et al., 2005a). Yet the genomic location of this QTL was remarkably consistent across studies, with the peak marker being RM252 (Gu et al., 2004, 2005a; Gu et al., 2005c), which is located at ~25 Mb on chromosome 4, or about 2 Mb from *Bh-a* and 6 Mb from *Phr1*. This strongly suggests that the accumulation of phenolic compounds in the rice hull has a pleiotropic effect on rice seed dormancy. It is therefore a real possibility that black hulls in rice were unconsciously selected against as humans selected for reduced seed dormancy. The transition to non-pigmented hulls would have brought about a concurrent reduction in seed dormancy, providing for the synchronous germination that would have given early agriculturalists additional control over rice cultivation.

SUMMARY

This study established the location and identity of the *Bh-a* gene controlling black hull coloration in rice, and in the process demonstrated a clear example of a two-gene epistatic interaction between physically linked genes. The results of this project led to the conclusion that a 22 bp deletion in an amino acid transporter gene on rice chromosome 4 is responsible for abolishing the appearance of black compounds in the mature rice hull. Since the mutation is at high frequency in both varietal groups of *O. sativa* and low frequency in *O. rufipogon*, this study concludes that the 22 bp deletion is associated with the domestication syndrome in rice. The previously reported polyphenol oxidase gene, *Phr1*, was also shown to be required for black hull coloration, although the divergence at this gene between *Indica* and *Japonica* suggests it was not under selection early during the rice domestication process. Future studies will investigate the origin of the 22 bp deletion and further characterize the molecular

and biochemical basis of polyphenol accumulation in rice hulls. Also, the *Bh-a* gene can now be isolated, and additional studies can establish whether the association between black hull coloration and seed dormancy in rice is due to linkage or pleiotropy.

APPENDIX

Supplemental Table 7.1: Accessions used for *Bh-a* association study

ID #	Subpopulation	Accession name	Origin	Hull Color	*Bh-a*	*PHR1*
1	*temperate japonica*	Agostano	Italy	Light	22bp	
2	*temperate japonica*	Aichi Ashahi	Japan	Light	WT	18bp del
3	*indica*	Ai-Chiao-Hong	China	Light	22bp	
4	*aus*	Arc 10177	India	Light	22bp	
5	*Group V*	Arc 10352	India	Light	22bp	
6	*aus*	Arc 7229	India	Light	22bp	
7	*tropical japonica*	Arias	Indonesia	Light	22bp	
8	*tropical japonica*	Asse Y Pung	Philippines	Light	22bp	
9	*temperate japonica*	Baber	India	Light	22bp	
10	*temperate japonica*	Baghlani Nangarhar	Afghanistan	Light	22bp	
11	*indica*	Baguamon 14	Bangladesh	Light	22bp	
12	*Group V*	Basmati	Pakistan	Light	22bp	
13	*aus*	Basmati 1	Pakistan	Light	22bp	
14	*Group V*	Basmati 217	India	Light	22bp	
15	*tropical japonica*	Beonjo	South Korea	Light	22bp	
16	*Group V*	Bico Branco	Brazil	Light	22bp	
17	*indica*	Binulawan	Philippines	Light	22bp	
18	*aus*	BJ 1	India	Light	22bp	
19	*aus*	Black Gora	India	Black	WT	WT
20	*tropical japonica*	Blue Rose	Louisiana	Light	22bp	
21	*indica*	Byakkoku Y 5006 Seln	Australia	Light	22bp	
22	*tropical japonica*	Caawa/Fortuna 6-103-15	Taiwan	Light	22bp	
23	*tropical japonica*	Canella De Ferro	Brazil	Light	22bp	
24	*tropical japonica*	Carolina Gold	Madagascar	Light	22bp	
25	*tropical japonica*	Carolina Gold	Madagascar	Light	22bp	
26	*tropical japonica*	Carolina Gold Sel	United States	Light	22bp	
27	*tropical japonica*	Chahora 144	Pakistan	Light	22bp	
27	*tropical japonica*	Chahora 144	Pakistan	Light	22bp	
28	*aus*	Champa Tong 54	Thailand	Light	22bp	
29	*indica*	Chau	Vietnam	Light	22bp	
30	*tropical japonica*	Chiem Chanh	Vietnam	Light	22bp	
31	*temperate japonica*	Chinese	China	Light	22bp	
32	*temperate japonica*	Chodongji	South Korea	Light	22bp	
33	*aus*	Chuan 4	Taiwan	Light	22bp	
34	*indica*	Co 25	India	Light	22bp	
35	*indica*	Co18	India	Light	22bp	
36	*tropical japonica*	Cs-M3	United States	Light	22bp	
37	*tropical japonica*	Cuba 65	Cuba	Light	22bp	
38	*tropical japonica*	DA 5	Bangladesh	Light	22bp	
39	*aus*	Da16	Bangladesh	Light	22bp	
40	*tropical japonica*	Dam	Thailand	Light	22bp	
41	*admix*	Darmali	Nepal	Light	22bp	
42	*temperate japonica*	Davao	Philippines	Light	22bp	
43	*indica*	Dee Geo Woo Gen	Taiwan	Light	22bp	
44	*aus*	Dhala Shaitta	Bangladesh	Light	22bp	
45	*Group V*	Dom-Sofid	Iran	Light	22bp	
46	*tropical japonica*	Dourado Agulha	Brazil	Light	22bp	
47	*indica*	Dourado Precoce	Brazil	Light	22bp	
48	*tropical japonica*	Dular	India	Light	22bp	
49	*aus*	DV85	Bangladesh	Light	22bp	
50	*aus*	DZ78	Bangladesh	Black	WT	WT
51	*temperate japonica*	Early Wataribune	Japan	Light	22bp	
52	*temperate japonica*	EH-1A-Chiu	Taiwan	Light	22bp	
53	*Group V*	Firooz	Iran	Light	22bp	
54	*tropical japonica*	Fortuna	United States	Light	22bp	
55	*temperate japonica*	Gerdeh	Iran	Light	22bp	
56	*temperate japonica*	Geumobyeo	South Korea	Light	22bp	
57	*indica*	Gharib	Iran	Light	22bp	

Supplemental Table 7.1 (continued)

58	*aus*	Ghati Kamma Nangarhar	Afghanistan	Light	22bp	
59	*tropical japonica*	Gogo Lempuk	Indonesia	Light	22bp	
60	*temperate japonica*	Gotak Gatik	Indonesia	Light	WT	18bp del
61	*indica*	Guan-Yin-Tsan	China	Light	22bp	
62	*temperate japonica*	Gyehwa 3	South Korea	Light	22bp	
63	*temperate japonica*	Haginomae Mochi	Japan	Light	22bp	
65	*tropical japonica*	Honduras	Honduras	Light	22bp	
66	*indica*	Hsia-Chioh-Keh-Tu	Taiwan	Light	22bp	
67	*temperate japonica*	Hu Lo Tao	China	Light	22bp	
68	*temperate japonica*	I-Geo-Tze	Taiwan	Light	22bp	
69	*tropical japonica*	IAC 25	Brazil	Light	22bp	
70	*tropical japonica*	Iguape Cateto	Haiti	Light	22bp	
71	*indica*	IR36	Philippines	Light	22bp	
72	*indica*	IR8	Philippines	Light	22bp	
73	*tropical japonica*	Irat 177	French Guiana	Light	22bp	
74	*indica*	Irga 409	Brazil	Light	22bp	
75	*tropical japonica*	Jambu	Indonesia	Light	22bp	
76	*indica*	Jaya	India	Light	22bp	
77	*indica*	JC149	India	Light	22bp	
78	*aus*	Jhona 349	India	Light	22bp	
79	*temperate japonica*	Jouiku 393G	Japan	Light	22bp	
80	*tropical japonica*	K 65	Suriname	Light	22bp	
81	*aus*	Kalamkati	India	Light	22bp	
82	*indica*	Kalukantha	Sri Lanka	Light	22bp	
83	*temperate japonica*	Kamenoo	Japan	Light	22bp	
84	*tropical japonica*	Kaniranga	Indonesia	Light	22bp	
85	*aus*	Kasalath	India	Light	WT	WT
86	*temperate japonica*	Kaw Luyoeng	Thailand	Light	22bp	
87	*tropical japonica*	Keriting Tingii	Indonesia	Light	22bp	
88	*aus*	Khao Gaew	Thailand	Light	22bp	
89	*tropical japonica*	Khao Hawm	Thailand	Light	22bp	
90	*indica*	Kiang-Chou-Chiu	Taiwan	Light	22bp	
91	*temperate japonica*	Kibi	Japan	Light	22bp	
92	*tropical japonica*	Kinastano	Philippines	Light	22bp	
93	*Group V*	Kitrana 508	Madagascar	Light	22bp	
94	*temperate japonica*	Koshihikari	Japan	Light	22bp	
95	*tropical japonica*	Kotobuki Mochi	Japan	Light	22bp	
96	*tropical japonica*	KU115	Thailand	Light	22bp	
97	*indica*	Kun-Min-Tsieh-Hunan	China	Light	22bp	
98	*tropical japonica*	L-202	United States	Light	22bp	
99	*tropical japonica*	Lac 23	Liberia	Light	22bp	
100	*tropical japonica*	Lacrosse	United States	Light	22bp	
101	*tropical japonica*	Lemont	United States	Light	22bp	
102	*temperate japonica*	Leung Pratew	Thailand	Light	22bp	
103	*temperate japonica*	Luk Takhar	Afghanistan	Light	22bp	
104	*temperate japonica*	Mansaku	Japan	Light	22bp	
105	*aus*	Mehr	Iran	Light	22bp	
106	*indica*	Ming Hui	China	Light	22bp	
107	*tropical japonica*	Miriti	Bangladesh	Light	22bp	
108	*tropical japonica*	Moroberekan	Guinea	Light	22bp	
109	*indica*	MTU9	India	Light	22bp	
110	*indica*	Mudgo	India	Light	22bp	
111	*tropical japonica*	N22	India	Light	22bp	
112	*Group V*	N12	India	Light	22bp	
113	*temperate japonica*	Norin 20	Japan	Light	Het	18bp del
114	*tropical japonica*	Nova	United States	Light	22bp	
115	*temperate japonica*	Npe 835	Pakistan	Light	22bp	
116	*tropical japonica*	NPE 844	Pakistan	Light	22bp	
117	*indica*	O-Luen-Cheung	Taiwan	Light	22bp	

Supplemental Table 7.1 (continued)

118	*temperate japonica*	Oro	Chile	Light	WT	18bp del
119	*indica*	Oryzica Llanos 5	Colombia	Light	22bp	
120	*tropical japonica*	Os6	Nigeria	Light	22bp	
121	*temperate japonica*	Ostiglia	Argentina	Light	22bp	
122	*tropical japonica*	Padi Kasalle	Indonesia	Light	22bp	
123	*indica*	Pagaiyahan	Taiwan	Light	22bp	
124	*indica*	Pankhari 203	India	Light	22bp	
125	*indica*	Pao-Tou-Hung	China	Light	22bp	
126	*indica*	Pappaku	Taiwan	Light	22bp	
127	*tropical japonica*	Patnai 23	India	Light	22bp	
128	*temperate japonica*	Pato De Gallinazo	Australia	Light	22bp	
129	*indica*	Peh-Kuh	Taiwan	Light	22bp	
130	*indica*	Peh-Kuh-Tsao-Tu	Taiwan	Light	22bp	
131	*aus*	Phudugey	Bhutan	Light	22bp	
132	*indica*	Rathuwee	Sri Lanka	Light	22bp	
133	*temperate japonica*	Rikuto Kemochi	Japan	Light	22bp	
134	*temperate japonica*	Romeo	Italy	Light	22bp	
135	*tropical japonica*	RT 1031-69	Zaire	Light	22bp	
136	*indica*	Rts12	Vietnam	Light	22bp	
137	*indica*	Rts14	Vietnam	Light	22bp	
138	*indica*	Rts4	Vietnam	Light	22bp	
139	*tropical japonica*	S4542A3-49B-2B12	United States	Light	22bp	
140	*tropical japonica*	Saturn	United States	Light	22bp	
141	*indica*	Seratoes Hari	Indonesia	Light	22bp	
142	*indica*	Shai-Kuh	China	Light	22bp	
143	*temperate japonica*	Shinriki	Japan	Light	22bp	
144	*temperate japonica*	Shoemed	United States	Light	22bp	
145	*indica*	Short Grain	Thailand	Light	22bp	
146	*indica*	Shuang-Chiang	Taiwan	Light	22bp	
147	*tropical japonica*	Sinampaga Selection	Philippines	Light	22bp	
148	*indica*	Sintane Diofor	Burkina Faso	Light	22bp	
149	*tropical japonica*	Sinaguing	Philippines	Light	22bp	
150	*tropical japonica*	Sultani	Egypt	Light	22bp	
151	*temperate japonica*	Suweon	Korea	Light	22bp	
151	*temperate japonica*	Suweon	Korea	N/A	22bp	
152	*aus*	T 1	India	Light	22bp	
153	*aus*	T26	India	Light	22bp	
154	*temperate japonica*	Ta Hung Ku	China	Light	WT	18bp del
155	*temperate japonica*	Ta Mao Tsao	China	Light	WT	18bp del
156	*indica*	Taichung Native 1	Taiwan	Light	22bp	
157	*temperate japonica*	Tainan Iku 487	Taiwan	Light	22bp	
158	*temperate japonica*	Taipei 309	Taiwan	Light	22bp	
159	*indica*	Tam Cau 9A	Vietnam	Light	22bp	
160	*Group V*	Tchampa	Iran	Light	22bp	
161	*indica*	Teqing	China	Light	22bp	
162	*indica*	Tkm6	India	Light	22bp	
163	*tropical japonica*	Taducan	Philippines	Light	22bp	
164	*tropical japonica*	Tondok	Indonesia	Light	22bp	
165	*tropical japonica*	Trembese	Indonesia	Light	22bp	
166	*indica*	Tsipala 421	Madagascar	Light	22bp	
167	*tropical japonica*	B6616A4-22-Bk-5-4	United States	Light	22bp	
168	*indica*	Vary Vato 462	Madagascar	Light	22bp	
169	*temperate japonica*	WC 6	China	Light	22bp	
170	*tropical japonica*	Wells	United States	Light	22bp	
171	*indica*	Zhe 733	China	Light	22bp	
172	*indica*	Zhenshan2	China	Light	22bp	
173	*temperate japonica*	Nipponbare	Japan	Light	22bp	
174	*tropical japonica*	Azucena	Philippines	Light	22bp	
175	*tropical japonica*	1021	Guatemala	Light	22bp	

Supplemental Table 7.1 (continued)

176	*tropical japonica*	583	Ecuador	Light	22bp	
177	*temperate japonica*	68-2	France	Light	22bp	
178	*aus*	Arc 6578	India	Light	22bp	
179	*temperate japonica*	Bellardone	France	Light	WT	18bp del
180	*temperate japonica*	Benllok	Peru	Light	22bp	
181	*temperate japonica*	Bergreis	Austria	Light	22bp	
182	*tropical japonica*	Blue Rose Supreme	United States	Light	22bp	
183	*tropical japonica*	Boa Vista	El Salvador	Light	22bp	
184	*temperate japonica*	Bombon	Spain	Light	22bp	
185	*temperate japonica*	British Honduras Creole	Belize	Light	22bp	
186	*temperate japonica*	Bul Zo	South Korea	Light	22bp	
187	*tropical japonica*	C57-5043	United States	Light	22bp	
188	*tropical japonica*	Coppocina	Bulgaria	Light	22bp	
189	*indica*	Criollo La Fria	Venezuela	Light	22bp	
190	*tropical japonica*	Delrex	United States	Light	22bp	
191	*Group V*	Dom Zard	Iran	Light	22bp	
192	*temperate japonica*	Erythroceros Hokkaido	Poland	Purple	WT	18bp del
193	*O. glaberrima*	Fossa Av	Burkina Faso	Light	Het	1bp del
194	*O. glaberrima*	HG24	Burkina Faso	Light	WT	1bp del
195	*tropical japonica*	Irat 13	Cote D'Ivoire	Light	22bp	
196	*indica*	JM70	Mali	Light	22bp	
197	*tropical japonica*	Kaukkyi Ani	Myanmar	Light	22bp	
198	*tropical japonica*	Leah	Bulgaria	Light	22bp	
199	*tropical japonica*	Mojito Colorado	Bolivia	Light	22bp	
200	*aus*	P 737	Pakistan	Light	22bp	
201	*tropical japonica*	Pate Blanc Mn 1	Cote D'Ivoire	Light	22bp	
202	*tropical japonica*	Pratao	Brazil	Light	22bp	
203	*indica*	Radin Ebos 33	Malaysia	Light	22bp	
204	*temperate japonica*	Razza 77	Italy	Light	22bp	
205	*temperate japonica*	Rinaldo Bersani	Italy	Light	22bp	
206	*indica*	Rojofotsy 738	Madagascar	Light	22bp	
207	*indica*	Sigadis	Indonesia	Light	22bp	
208	*indica*	Slo 17	India	Light	22bp	
209	*indica*	Tchibanga	Gabon	Light	22bp	
210	*aus*	Thavalu	Sri Lanka	Light	22bp	
211	*tropical japonica*	Tokyo Shino Mochi	Japan	Light	22bp	
212	*tropical japonica*	Wc 2810	Micronesia	Light	22bp	
213	*tropical japonica*	Wc 3397	Jamaica	Light	22bp	
214	*tropical japonica*	Wc 4419	Honduras	Light	22bp	
215	*tropical japonica*	Wc 4443	Bolivia	Light	22bp	
216	*temperate japonica*	Yabani Montakhab 7	Egypt	Light	22bp	
217	*temperate japonica*	Yrl-1	Australia	Light	22bp	
218	*tropical japonica*	Pi 298967-1	Australia	Light	22bp	
219	*temperate japonica*	Nucleoryza	Hungary	Light	WT	18bp del
220	*temperate japonica*	Azerbaidjanica	Azerbaijan	Light	22bp	
221	*Group V*	Sadri Belyi	Azerbaijan	Light	22bp	
222	*indica*	Paraiba Chines Nova	Brazil	Light	22bp	
223	*tropical japonica*	Priano Guaira	Brazil	Light	22bp	
224	*admix*	Karabaschak	Bulgaria	Light	WT	WT
225	*temperate japonica*	Biser 1	Bulgaria	Light	22bp	
226	*tropical japonica*	Irat 44	Burkina Faso	Light	22bp	
227	*aus*	Riz Local	Burkina Faso	Light	22bp	
228	*aus*	Ca 902/B/2/1	Chad	Light	22bp	
229	*temperate japonica*	Niquen	Chile	Light	WT	18bp del
230	*temperate japonica*	Oro	Chile	Light	WT	18bp del
231	*indica*	Hunan Early Dwarf No3	China	Light	22bp	
232	*temperate japonica*	Shangyu 394	China	Light	22bp	
233	*temperate japonica*	Sung Liao 2	China	Light	22bp	
234	*indica*	Aijiaonante	China	Light	22bp	

Supplemental Table 7.1 (continued)

235	*indica*	Sze Guen Zim	China	Light	22bp	
236	*tropical japonica*	Wc 521	China	Light	22bp	
237	*temperate japonica*	Estrela	Colombia	Light	22bp	
238	*tropical japonica*	WAB56-104	Cote D'Ivoire	Light	22bp	
239	*tropical japonica*	WAB 502-13-4-1	Cote D'Ivoire	Light	22bp	
240	*tropical japonica*	Wab 501-11-5-1	Cote D'Ivoire	Light	22bp	
241	*indica*	Ecia76-S89-1	Cuba	Light	22bp	
242	*tropical japonica*	27	Dominican Republic	Light	22bp	
243	*temperate japonica*	Tropical Rice	Ecuador	Light	22bp	
244	*temperate japonica*	Arabi	Egypt	Light	22bp	
245	*temperate japonica*	Sab Ini	Egypt	Light	22bp	
246	*Group V*	Saraya	Fiji	Light	22bp	
247	*temperate japonica*	Desvauxii	Former Soviet Union	Light	WT	18bp del
248	*temperate japonica*	Caucasica	Former Soviet Union	Light	22bp	
249	*indica*	Pirinae 69	Former Yugoslavia	Light	22bp	
250	*temperate japonica*	Bulgare	France	Light	WT	18bp del
251	*tropical japonica*	H256-76-1-1-1	Argentina	Light	22bp	
252	*indica*	Djimoron	Guinea	Light	22bp	
253	*temperate japonica*	Guineandao	Guinea	Light	22bp	
254	*indica*	Hon Chim	Hong Kong	Light	22bp	
255	*indica*	Pai Hok Glutinous	Hong Kong	Light	22bp	
256	*temperate japonica*	Romanica	Hungary	Light	WT	18bp del
257	*temperate japonica*	Agusita	Hungary	Light	WT	18bp del
258	*tropical japonica*	Tia Bura	Indonesia	Light	22bp	
259	*indica*	Sadri Tor Misri	Iran	Light	22bp	
260	*Group V*	205	Iran	Light	22bp	
261	*aus*	Shim Balte	Iraq	Light	22bp	
262	*admix*	Halwa Gose Red	Iraq	Light	WT	WT
263	*temperate japonica*	Maratelli	Italy	Light	22bp	
264	*temperate japonica*	Baldo	Italy	Light	22bp	
265	*temperate japonica*	Vialone	Italy	Light	22bp	
266	*temperate japonica*	Hiderisirazu	Japan	Light	22bp	
267	*temperate japonica*	Hatsunishiki	Japan	Light	Het	18bp del
268	*temperate japonica*	Vavilovi	Kazakhstan	Light	WT	18bp del
269	*indica*	Sundensis	Kazakhstan	Light	22bp	
270	*temperate japonica*	Osogovka	Macedonia	Light	22bp	
270	*temperate japonica*	Osogovka	Macedonia	Light	22bp	
271	*temperate japonica*	M Blatec	Macedonia	Light	22bp	
272	*indica*	923	Madagascar	Light	22bp	
273	*tropical japonica*	Varyla	Madagascar	Light	22bp	
274	*tropical japonica*	Padi Pagalong	Malaysia	Light	22bp	
275	*temperate japonica*	Sri Malaysia Dua	Malaysia	Light	WT	18bp del
276	*aus*	Kaukau	Mali	Light	22bp	
277	*temperate japonica*	Gambiaka Sebela	Mali	Light	22bp	
278	*temperate japonica*	C1-6-5-3	Mexico	Light	22bp	
279	*temperate japonica*	Kon Suito	Mongolia	Light	WT	18bp del
280	*tropical japonica*	Saku	Mongolia	Light	22bp	
281	*temperate japonica*	Patna	Morocco	Light	22bp	
282	*temperate japonica*	Triomphe Du Maroc	Morocco	Light	WT	18bp del
283	*temperate japonica*	Chibica	Mozambique	Light	22bp	
284	*indica*	IR-44595	Nepal	Light	22bp	
285	*tropical japonica*	Tox 782-20-1	Nigeria	Light	22bp	
286	*tropical japonica*	Iita 135	Nigeria	Light	22bp	
287	*temperate japonica*	Zerawchanica Karatalski	Poland	Light	22bp	
288	*temperate japonica*	Italica Carolina	Poland	Light	22bp	
289	*temperate japonica*	Lusitano	Portugal	Light	WT	18bp del
290	*temperate japonica*	Amposta	Puerto Rico	Light	22bp	
291	*temperate japonica*	Toploea 70/76	Romania	Light	22bp	
291	*temperate japonica*	Toploea 70/76	Romania	Light	22bp	

Supplemental Table 7.1 (continued)

292	*temperate japonica*	Stegaru 65	Romania	Light	22bp	
293	*aus*	Tog 7178	Senegal	Light	22bp	
294	*indica*	SL 22-613	Sierra Leone	Light	22bp	
295	*temperate japonica*	Bombilla	Spain	Light	22bp	
296	*temperate japonica*	Dosel	Spain	Light	22bp	
297	*temperate japonica*	Bahia	Spain	Light	22bp	
298	*indica*	Ld 24	Sri Lanka	Light	22bp	
299	*indica*	SML 242	Suriname	Light	22bp	
300	*temperate japonica*	Sml Kapuri	Suriname	Light	22bp	
301	*temperate japonica*	Melanotrix	Tajikistan	Light	22bp	
302	*temperate japonica*	Wir 3039	Tajikistan	Light	22bp	
303	*temperate japonica*	Kihogo	Tanzania	Light	22bp	
304	*indica*	519	Uruguay	Light	22bp	
305	*temperate japonica*	Doble Carolina Rinaldo Barsani	Uruguay	Light	22bp	
306	*temperate japonica*	Wir 3764	Uzbekistan	Light	WT	18bp del
307	*temperate japonica*	Uzbekskij 2	Uzbekistan	Light	22bp	
308	*tropical japonica*	Llanero 501	Venezuela	Light	22bp	
309	*tropical japonica*	Manzano	Zaire	Light	22bp	
310	*tropical japonica*	R 101	Zaire	Light	22bp	
311	*temperate japonica*	56-122-23	Thailand	Light	22bp	
312	*aus*	Aswina 330	Bangladesh	Light	WT	WT
313	*indica*	Br24	Bangladesh	Light	22bp	
314	*aus*	Ctg 1516	Bangladesh	Light	22bp	
315	*indica*	Dawebyan	Myanmar	Light	22bp	
316	*aus*	Dd 62	Bangladesh	Light	22bp	
317	*aus*	Dj 123	Bangladesh	Light	22bp	
318	*aus*	Dj 24	Bangladesh	Light	22bp	
319	*aus*	DK12	Bangladesh	Light	22bp	
320	*aus*	Dm 43	Bangladesh	Light	22bp	
321	*aus*	Dm 56	Bangladesh	Light	22bp	
322	*aus*	DM 59	Bangladesh	Light	WT	WT
323	*aus*	Dnj 140	Bangladesh	Light	22bp	
324	*aus*	Dv 123	Bangladesh	Light	22bp	
325	*indica*	Emata A 16-34	Myanmar	Light	22bp	
326	*aus*	Ghorbhai	Bangladesh	Light	22bp	
327	*aus*	Goria	Bangladesh	Black	WT	WT
328	*aus*	Jamir	Bangladesh	Light	22bp	
329	*aus*	Kachilon	Bangladesh	Light	22bp	
330	*aus*	Khao Pahk Maw	Thailand	Light	22bp	
331	*aus*	Khao Tot Long 227	Thailand	Light	22bp	
332	*indica*	Kpf-16	Bangladesh	Light	22bp	
333	*temperate japonica*	Leuang Hawn	Thailand	Light	22bp	
334	*temperate japonica*	Lomello	Thailand	Light	22bp	
335	*tropical japonica*	Okshitmayin	Myanmar	Light	22bp	
336	*aus*	Paung Malaung	Myanmar	Light	22bp	
337	*indica*	Sabharaj	Bangladesh	Light	22bp	
338	*temperate japonica*	Sitpwa	Myanmar	Light	22bp	
339	*indica*	Yodanya	Myanmar	Light	22bp	
340	*Group V*	Berenj	Afghanistan	Light	22bp	
341	*aus*	Shirkati	Afghanistan	Light	22bp	
342	*tropical japonica*	Cenit	Argentina	Light	22bp	
343	*temperate japonica*	Victoria F.A.	Argentina	Light	22bp	
344	*Group V*	Habiganj Boro 6	Bangladesh	Light	22bp	
345	*aus*	Dz 193	Bangladesh	Light	22bp	
346	*aus*	Karkati 87	Bangladesh	Light	22bp	
347	*tropical japonica*	Creole	Belize	Light	22bp	
348	*indica*	China 1039	China	Light	22bp	
349	*indica*	Chang Ch'Sang Hsu Tao	China	Light	22bp	
350	*tropical japonica*	Ligerito	Colombia	Light	22bp	

Supplemental Table 7.1 (continued)

351	*temperate japonica*	68-2	France	Light	22bp	
352	*tropical japonica*	Guatemala 1021	Guatemala	Light	22bp	
353	*aus*	Arc 10376	India	Light	22bp	
354	*indica*	Bala	India	Light	22bp	
355	*temperate japonica*	Asd 1	India	Light	WT	18bp del
356	*indica*	Jc 117	India	Light	22bp	
357	*aus*	9524	India	Light	22bp	
358	*Group V*	ARC 10086	India	Light	22bp	
359	*aus*	Surjamkuhi	India	Light	22bp	
360	*aus*	Ptb 30	India	Light	22bp	
361	*temperate japonica*	F.R. 13A	India	Light	22bp	
362	*tropical japonica*	Jamaica 3	Jamaica	Light	22bp	
363	*temperate japonica*	Edomen Scented	Japan	Light	22bp	
363	*temperate japonica*	Edomen Scented	Japan	Light	22bp	
364	*tropical japonica*	Rikuto Norin 21	Japan	Light	22bp	
365	*temperate japonica*	Shirogane	Japan	Light	22bp	
366	*temperate japonica*	Kiuki No. 46	Japan	Light	22bp	
366	*temperate japonica*	Kiuki No. 46	Japan	Light	22bp	
367	*tropical japonica*	Sanbyang-Daeme	Korea	Light	22bp	
368	*temperate japonica*	Deokjeokjodo	Korea	Light	22bp	
369	*aus*	Sathi	Pakistan	Light	22bp	
370	*aus*	Coarse	Pakistan	Light	22bp	
371	*aus*	Santhi Sufaid	Pakistan	Light	22bp	
372	*aus*	Sufaid	Pakistan	Light	22bp	
373	*Group V*	Lambayeque 1	Peru	Light	22bp	
374	*temperate japonica*	Benllok	Peru	Light	22bp	
375	*tropical japonica*	Upland	PONAPE ISLAND	Light	22bp	
376	*temperate japonica*	Breviaristata	Portugal	Light	22bp	
377	*tropical japonica*	Pr 304	Puerto Rico	Light	22bp	
378	*aus*	Kalubala Vee	Sri Lanka	Black	WT	WT
379	*tropical japonica*	Wanica	Suriname	Light	22bp	
380	*temperate japonica*	Tainan-Iku No. 512	Taiwan	Light	WT	18bp del
381	*tropical japonica*	325	Taiwan	Light	22bp	
382	*tropical japonica*	KU 115	Thailand	Light	22bp	
383	*temperate japonica*	Coll 2712	TURKEY	Light	22bp	
384	*tropical japonica*	318	TURKEY	Light	22bp	
385	*indica*	Nira	United States	Light	22bp	
386	*tropical japonica*	Palmyra	United States	Light	22bp	
387	*temperate japonica*	M-202	United States	Light	22bp	
388	*temperate japonica*	Nortai	United States	Light	22bp	
389	*tropical japonica*	Ci 11011	United States	Purple	22bp	WT
390	*indica*	Cl 11026	United States	Light	22bp	
391	*tropical japonica*	Della	United States	Light	22bp	
392	*tropical japonica*	Edith	United States	Light	22bp	
393	*indica*	LA 110	United States	Light	22bp	
394	*tropical japonica*	Lady Wright Seln	United States	Light	22bp	
395	*tropical japonica*	Os 6 (Wc 10296)	Zaire	Light	22bp	
396	*tropical japonica*	Cocodrie	United States	Light	22bp	
397	*tropical japonica*	Cybonnet	United States	Light	22bp	
398	*indica*	93-11		Light	22bp	
399	*tropical japonica*	Spring	United States	Light	22bp	
400	*indica*	Yang Dao 6		Light	22bp	
616	*indica*	Rt0034	United States	Light	22bp	
617	*tropical japonica*	Mcr010277	United States	Light	22bp	
618	*tropical japonica*	Pecos	United States	Light	22bp	
619	*tropical japonica*	Rosemont	United States	Light	22bp	
619	*tropical japonica*	Rosemont	United States	Light	22bp	
622	*tropical japonica*	Bengal		Light	22bp	
627	*tropical japonica*	Early (Farmbuster)	United States	Light	22bp	

Supplemental Table 7.1 (continued)

628	*tropical japonica*	Jefferson	United States	Light	22bp	
629	*tropical japonica*	Panda	United States	Light	22bp	
630	*tropical japonica*	Saber	United States	Light	22bp	
631	*temperate japonica*	Dragon Eyeball 100	China	Purple	22bp	1bp ins
632	*tropical japonica*	Francis	United States	Light	22bp	
633	*indica*	Jing 185 185-7	China	Light	22bp	
634	*indica*	4484 1693	China	Light	22bp	
635	*tropical japonica*	Azucena	Philippines	Light	22bp	
636	*indica*	Sadu Cho	Korea	Light	22bp	
637	*tropical japonica*	N22	India	Light	22bp	
638	*tropical japonica*	Moroberekan	Guinea	Light	22bp	
639	*temperate japonica*	Nipponbare	Japan	Light	22bp	
640	*Group V*	Dom-Sofid	Iran	Light	22bp	
641	*temperate japonica*	Tainung 67	Taiwan	Light	WT	18bp del
642	*indica*	Zhenshan 97B	China	Light	22bp	
643	*indica*	Minghui 63	China	Light	22bp	
644	*indica*	Ir64	Philippines	Light	22bp	
645	*temperate japonica*	M202	United States	Light	22bp	
646	*indica*	Swarna	Iran	Light	22bp	
647	*tropical japonica*	Cypress	United States	Light	22bp	
649	*aus*	Fr13A	India	Light	WT	WT
650	*admix*	Aswina	Bangladesh	Light	WT	WT
651	*aus*	Dular	India	Light	22bp	
654	*indica*	Pokkali		Light	22bp	
764	*temperate japonica*	Hwayheong		Light	22bp	
401_B	*O. rufipogon*		India	Black	WT	
402_B	*O. rufipogon*		India	Black	WT	
404_A	*O. rufipogon*		India	Black	WT	
405_A	*O. rufipogon*		India	Black	WT	
407_C	*O. rufipogon*		Myanmar	Light	22bp	
408_B	*O. rufipogon*		Myanmar	Light	22bp	
413_A	*O. rufipogon*		India	Black	WT	
414_C	*O. rufipogon*		India	Black	WT	
415_B	*O. rufipogon*		India	Black	WT	
416_A	*O. rufipogon*		Thailand	Light	22bp	
417_B	*O. rufipogon*		Indonesia	Black	WT	
420_A	*O. rufipogon*		Laos	Black	Het	
427_C	*O. rufipogon*		China	Black	WT	
428_A	*O. rufipogon*		China	Black	Het	
429_A	*O. rufipogon*		China	Black	Het	
430_B	*O. rufipogon*		China	Black	Het	
431_A	*O. rufipogon*		China	Black	WT	
432_B	*O. rufipogon*		Thailand	Black	WT	
435_C	*O. rufipogon*		Thailand	Black	WT	
437_B	*O. rufipogon*		India	Black	WT	
438_B	*O. rufipogon*		India	Black	WT	
440_B	*O. rufipogon*		Bangladesh	Black	WT	
442_A	*O. rufipogon*		Nepal	Black	WT	
443_B	*O. rufipogon*		Nepal	Black	WT	
444_A	*O. rufipogon*		Nepal	Black	WT	
445B	*O. rufipogon*		Nepal	Black	WT	
446_A	*O. rufipogon*		Nepal	Black	WT	
449_A	*O. rufipogon*		Myanmar	Black	WT	
450_A	*O. rufipogon*		China	Black	WT	
451_B	*O. rufipogon*		India	Light	22bp	
452_A	*O. rufipogon*		Taiwan	Black	WT	
453_C	*O. rufipogon*		Bangladesh	Black	Het	
454_A	*O. rufipogon*		China	Black	WT	
456_A	*O. rufipogon*		China	Black	WT	

Supplemental Table 7.1 (continued)

457_B	*O. rufipogon*		Bangladesh	Light	22bp	
458_B	*O. rufipogon*		Bangladesh	Light	22bp	
460_C	*O. rufipogon*		India	Black	Het	
461_A	*O. rufipogon*		China	Black	Het	
462A	*O. rufipogon*		India	Black	WT	
463_B	*O. rufipogon*		Sri Lanka	Black	WT	
465_A	*O. rufipogon*		China	Black	WT	
467_A	*O. rufipogon*		China	Black	WT	
469_B	*O. rufipogon*		China	Black	WT	
471_C	*O. rufipogon*		China	Black	Het	
472_B	*O. rufipogon*		China	Black	WT	
473_A	*O. rufipogon*		Thailand	Black	WT	
474_C	*O. rufipogon*		Thailand	Black	WT	
475_B	*O. rufipogon*		Thailand	Black	WT	
477_A	*O. rufipogon*		China	Light	22bp	
479_A	*O. rufipogon*		Indonesia	Black	WT	
481_C	*O. rufipogon*		India	Black	WT	
482_A	*O. rufipogon*		India	Black	WT	
483_C	*O. rufipogon*		Thailand	Black	WT	
484_B	*O. rufipogon*		Thailand	Black	WT	
487_C	*O. rufipogon*		Sri Lanka	Black	WT	
488_B	*O. rufipogon*		Malaysia	Black	WT	
490_A	*O. rufipogon*		Indonesia	Black	WT	
491_A	*O. rufipogon*		Philippines	Black	WT	
492B	*O. rufipogon*		Cambodia	Black	WT	
493_A	*O. rufipogon*		Nepal	Black	WT	
494_a	*O. rufipogon*		India	Black	WT	
495_a	*O. rufipogon*		Cambodia	Black	WT	
496_A	*O. rufipogon*		Cambodia	Black	WT	
498_A	*O. rufipogon*		Cambodia	Black	WT	
499_B	*O. rufipogon*		Thailand	Black	Het	
500_A	*O. rufipogon*		Thailand	Black	WT	
501_B	*O. rufipogon*		Thailand	Black	WT	
503_C	*O. rufipogon*		Thailand	Black	Het	
505_A	*O. rufipogon*		Thailand	Light	22bp	
506_A	*O. rufipogon*		Bangladesh	Black	WT	
509A	*O. rufipogon*		Bangladesh	Black	WT	
514_C	*O. rufipogon*		Indonesia	Black	WT	
518_A	*O. rufipogon*		India	Black	WT	
523_A	*O. rufipogon*		Laos	Black	WT	
538A	*O. rufipogon*		Vietnam	Black	WT	
540_B	*O. rufipogon*		Vietnam	Black	WT	
549_A	*O. rufipogon*		India	Black	WT	
551_C	*O. rufipogon*		Taiwan	Light	22bp	
553_B	*O. rufipogon*		Myanmar	Black	Het	
555_B	*O. rufipogon*		India	Black	WT	
556_A	*O. rufipogon*		Myanmar	Black	WT	
557_B	*O. rufipogon*		Indonesia	Black	Het	
558_A	*O. rufipogon*		unknown	Black	WT	
559_C	*O. rufipogon*		unknown	Black	Het	
568_A	*O. rufipogon*		Papua New Guinea	Black	WT	
579_C	*O. rufipogon*		Papua New Guinea	Black	WT	
592_B	*O. rufipogon*		India	Light	22bp	
599_A	*O. rufipogon*		India	Black	WT	
600_B	*O. rufipogon*		Malaysia	Light	22bp	
602_A	*O. rufipogon*		India	Black	WT	
604_C	*O. rufipogon*		Taiwan	Light	22bp	
605_C	*O. rufipogon*		Thailand	Light	22bp	
607_A	*O. rufipogon*			Black	WT	

Supplemental Table 7.1 (continued)

608_B	*O. rufipogon*			Black	WT	
655B	*O. rufipogon*		India	Black	WT	
656_B	*O. rufipogon*			Light	22bp	
657A	*O. rufipogon*		India	Black	WT	
665_C	*O. rufipogon*		Myanmar	Light	22bp	
666B	*O. rufipogon*		India	Black	WT	
669_C	*O. rufipogon*		Taiwan	Black	WT	
669C	*O. rufipogon*			Black	WT	
670C	*O. rufipogon*		Taiwan	Light	22bp	
673_A	*O. rufipogon*		Taiwan	Light	22bp	
676A	*O. rufipogon*		Taiwan	Light	22bp	
678C	*O. rufipogon*		India	Black	WT	
679_B	*O. rufipogon*		India	Black	WT	
681_C	*O. rufipogon*		India	Black	WT	
682_C	*O. rufipogon*		Thailand	Black	WT	
683_A	*O. rufipogon*		Cambodia	Black	WT	
686_C	*O. rufipogon*		Myanmar	Black	Het	
687_A	*O. rufipogon*		Taiwan	Light	22bp	
688_A	*O. rufipogon*		Taiwan	Black	WT	
691_A	*O. rufipogon*		India	Black	WT	
693_A	*O. rufipogon*		Cambodia	Black	WT	
694_A	*O. rufipogon*		Taiwan	Black	WT	
695_B	*O. rufipogon*		Sri Lanka	Black	WT	
702_C	*O. rufipogon*		China	Black	WT	
707_B	*O. rufipogon*		Bangladesh	Black	WT	
708_A	*O. rufipogon*		Bangladesh	Black	WT	
710_B	*O. rufipogon*		Bangladesh	Black	WT	
711_A	*O. rufipogon*		Bangladesh	Black	WT	
712_B	*O. rufipogon*		India	Black	WT	
715_B	*O. rufipogon*		Thailand	Black	WT	
716_B	*O. rufipogon*		Thailand	Black	WT	
717_B	*O. rufipogon*		Thailand	Black	WT	
719_A	*O. rufipogon*		India	Black	WT	
720_A	*O. rufipogon*		India	Black	WT	
721_C	*O. rufipogon*		India	Black	WT	
722_A	*O. rufipogon*		China	Black	WT	
723_B	*O. rufipogon*		China	Black	WT	
725_A	*O. rufipogon*		India	Black	WT	
727_B	*O. rufipogon*		Thailand	Black	WT	
733_B	*O. rufipogon*		Sri Lanka	Black	WT	
735_C	*O. rufipogon*		Myanmar	Black	WT	
736_B	*O. rufipogon*			Black	Het	
737_B	*O. rufipogon*		Thailand	Black	WT	
738_B	*O. rufipogon*		Thailand	Black	WT	
741_A	*O. rufipogon*		India	Black	WT	
743_C	*O. rufipogon*		NPL	Black	WT	
746_C	*O. rufipogon*		Cambodia	Black	WT	
747_C	*O. rufipogon*		Cambodia	Black	WT	
748_C	*O. rufipogon*		Thailand	Black	WT	
749_C	*O. rufipogon*		Thailand	Black	WT	
751_C	*O. rufipogon*		Bangladesh	Light	22bp	
757_A	*O. rufipogon*		Laos	Black	WT	
759_A	*O. rufipogon*		Cambodia	Black	WT	
760_A	*O. rufipogon*		Myanmar	Black	WT	
762_C	*O. rufipogon*		Myanmar	Black	WT	

REFERENCES

Aladejana F.O. (2000) Botanical and cytogenetic studies of some indegenous species in the A-genome complex of the genus Oryza Linn. Department of Botany. Obafemi Awolowo University, Ile-Ife

Alexandrov N., Brover V., Freidin S., Troukhan M., Tatarinova T., Zhang H., Swaller T., Lu Y.-P., Bouck J., Flavell R., Feldmann K. (2009) Insights into corn genes derived from large-scale cDNA sequencing. Plant Molecular Biology 69:179-194

Bewley J.D., Black M. (1982) Physiology and biochemistry of seeds in relation to germination 2. Viability, dormancy, and environmental control. Springer-Verlag, Berlin

Cai, Cai H., Morishima, Morishima H. (2002) QTL clusters reflect character associations in wild and cultivated rice. Theoretical and Applied Genetics 104:1217-1228

Causse M.A., Fulton T.M., Cho Y.G., Ahn S.N., Chunwongse J., Wu K., Xiao J., Yu Z., Ronald P.C., Harrington S.E., Second G., McCouch S.R., Tanksley S.D. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251-1274

Chang, T.T. & Bardenas, E.A. 1965. The morphology and varietal characters of the rice plant. Technical Bulletin 4, IRRI, Philippines

Chao L.F. (1928) Linkage studies in rice. Genetics 13:133-169

Chen M., Presting G., Barbazuk W.B., Goicoechea J.L., Blackmon B., Fang G., Kim H., Frisch D., Yu Y., Sun S., Higingbottom S., Phimphilai J., Phimphilai D., Thurmond S., Gaudette B., Li P., Liu J., Hatfield J., Main D., Farrar K., Henderson C., Barnett L., Costa R., Williams B., Walser S., Atkins M., Hall C., Budiman M.A., Tomkins J.P., Luo M., Bancroft I., Salse J., Regad F.,

Mohapatra T., Singh N.K., Tyagi A.K., Soderlund C., Dean R.A., Wing R.A. (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537-545

Debeaujon I., Leon-Kloosterziel K.M., Koornneef M. (2000) Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. Plant Physiology 122:403-414

Debeaujon I., Lepiniec L., Pourcel L., Routaboul J.M. (2007) Seed coat development and dormancy. In: Bradford K.J.N., H. (ed) Seed Development, Dormancy and Germination. Blackwell, Oxford, pp 25-49

DePauw R.M., McCaig T.N. (1983) Recombining dormancy and white seed color in a spring wheat cross. Canadian Journal of Plant Science 63:581-589

Doebley J.F., Gaut B.S., Smith B.D. (2006) The molecular genetics of crop domestication. Cell 127:1309-1321

Doi K., Izawa T., Fuse T., Yamanouchi U., Kubo T., Shimatani Z., Yano M., Yoshimura A. (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT*-like gene expression independently of *Hd1*. Genes & Development 18:926-936

Fan C., Xing Y., Mao H., Lu T., Han B., Xu C., Li X., Zhang Q. (2006) *GS3* , a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and Applied Genetics 112:1164-1171

Finch-Savage W.E., Leubner-Metzger G. (2006) Seed dormancy and the control of germination. New Phytologist 171:501-523

Finkelstein R., Reeves W., Ariizumi T., Steber C. (2008) Molecular aspects of seed dormancy. Annual Review of Plant Biology 59:387-415

Furukawa T., Maekawa M., Oki T., Suda I., Iida S., Shimada H., Takamure I., Kadowaki K.-i. (2007) The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. The Plant Journal 49:91-102

Gepts P. (2004) Crop domestication as a long-term selection experiment. Plant Breeding Reviews 24:1-44

Gfeller F., Svejda F. (1960) Inheritance of post-harvest seed dormancy and kernel colour in spring wheat lines. Canadian Journal of Plant Science 40:1-6

Groos C., Gay G., Perretant M.R., Gervais L., Bernard M., Dedryver F., Charmet G. (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. Theoretical and Applied Genetics 104:39-47

Gu X.-Y., Kianian S.F., Foley M.E. (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). Genetics 166:1503-1516

Gu X.-Y., Kianian S.F., Foley M.E. (2005a) Phenotypic selection for dormancy introduced a set of adaptive haplotypes from weedy into cultivated rice. Genetics 171:695-704

Gu X.-Y., Kianian S.F., Foley M.E. (2005b) Seed dormancy imposed by covering tissues interrelates to shattering and seed morphological characteristics in weedy rice. Crop Science 45:948-955

Gu X.-Y., Kianian S.F., Hareland G.A., Hoffer B.L., Foley M.E. (2005c) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theoretical and Applied Genetics 110:1108-1118

Gu X.Y., Kianian S.F., Foley M.E. (2006) Isolation of three dormancy QTLs as Mendelian factors in rice. Heredity 96:93-99

Harushima Y., Yano M., Shomura A., Sato M., Shimano T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A., Kajiya H., Huang N., Yamamoto K., Nagamura Y., Kurata N., Khush G.S., Sasaki T. (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479-494

Hu F.Y., Tao D.Y., Sacks E., Fu B.Y., Xu P., Li J., Yang Y., McNally K., Khush G.S., Paterson A.H., Li Z.-K. (2003) Convergent evolution of perenniality in rice and sorghum. Proceedings of the National Academy of Sciences of the United States of America 100:4050-4054

Jena K.K., Khush G.S. (1990) Introgression of genes from *Oryza officinalis* Well ex Watt to cultivated rice, *O. sativa* L. Theoretical and Applied Genetics 80:737-745

Ji H.-S., Chu S.-H., Jiang W., Cho Y.-I., Hahn J.-H., Eun M.-Y., McCouch S.R., Koh H.-J. (2006) Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. Genetics 173:995-1005

Jin J., Huang W., Gao J.-P., Yang J., Shi M., Zhu M.-Z., Luo D., Lin H.-X. (2008) Genetic control of rice plant architecture under domestication. Nature Genetics 40:1365-1369

Kasahara Y. (1941) Varietal identification of husked rice by means of phenol and paracresol staining. Proceedings of the Crop Science Society of Japan 13:81-88

Khan F.N., Strand E.A. (1977) Investigations into the genetics of kernel color and dormancy in wheat (*Triticum aestivum* L.). Meld Nor Landbrukshogsk 56:1-12

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant and Cell Physiology 43:1096-1105

Kondo M., Kasahara Y. (1940) On the discrimination of rice varieties by means of colouring reaction of hulled kernels to phenol-fuchshin. Proceedings of the Crop Science Society of Japan 12:122-128

Konishi S., Izawa T., Lin S.Y., Ebana K., Fukuta Y., Sasaki T., Yano M. (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392-1396

Kovach M.J., Sweeney M.T., McCouch S.R. (2007) New insights into the history of rice domestication. Trends in Genetics 23:578-587

Kuang H.H., Tu D.S., Chang Y.H. (1946) Linkage studies of awn in cultivated rice (*Oryza sativa* L.). Journal of Genetics 47:249-259

Kuriyama H., Kudo M. (1967) Complementary genes *Ph* and *Bh* controlling ripening-black coloration of rice hulls and their geographical distribution. Japanese Journal of Breeding 17:13-19

Li C., Zhou A., Sang T. (2006) Rice domestication by reducing shattering. Science 311:1936-1939

Li L., Steffens J. (2002) Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance. Planta 215:239-247

Lin S.Y., Nagamura Y., Kurata N., Yano M., Minobe Y., Sasaki T. (1994) DNA markers tightly linked to genes, *Ph*, *Alk*, and *Rc*. Rice Genetics Newsletter 11:108-109

Lin Z., Griffith M., Li X., Zhu Z., Tan L., Fu Y., Zhang W., Wang X., Xie D., Sun C. (2007) Origin of seed shattering in rice (*Oryza sativa* L.). Planta 226:11-20

Maekawa M. (1982) Studies on genetical differences between distantly related rice varieties. Memoirs of the Faculty of Agriculture, Hokkaido University 13:146-177

Maekawa M. (1984) Geographical distribution of the genes for black hull coloration. Rice Genetics Newsletter 1:104

Mayer A.M. (2006) Polyphenol oxidases in plants and fungi: Going places? A review. Phytochemistry 67:2318-2331

McCouch S.R., Kochert G., Yu Z.H., Wang Z.Y., Khush G.S., Coffman W.R., Tanksley S.D. (1988) Molecular mapping of rice chromosomes. Theoretical and Applied Genetics 76:815-829

McCouch S.R., Kovach M.J., Sweeney M.T., Semon M. (2010) Gene flow and genetic isolation in crop evolution. In: Gepts P. (ed) Proceedings of the Harlan II Symposium: Biodiversity in Agriculture: Domestication, Evolution, & Sustainability. Cambridge University Press, UC Davis, CA

Melo G.A., Shimizu M.M., Mazzafera P. (2006) Polyphenoloxidase activity in coffee leaves and its role in resistance against the coffee leaf miner and coffee leaf rust. Phytochemistry 67:277-285

Mitra S.K., Ganguli P.M. (1936) Inheritance of inner glume colour in rice. Indian Journal of Agricultural Science 7:126-133

Moringa T., Nagamatu T., Kawahara E. (1943) New linkage relation in rice. Japanese Journal of Genetics 19:206-208

Nagao S., Takahashi M. (1952) Genetical studies on rice plant XIV. The order and distance of some genes belonging to PL-linkage group. Japanese Journal of Breeding 1:232-240

Nagao S., Takahashi M. (1954) Genetical studies on rice plant. XVI. Some genes responsible for yellow, brown and black color of glume. Japanese Journal of Breeding 4:25-30

Nwokeocha C.C. (1998) Botanical and cytogenetic studies in genus *Oryza* Linn. Obafemi Awolowo University, Ile-Ife

Oka H.I. (1953) Phylogenetic differentation of the cultivated rice plant 1. Variations in respective characters and their combinations in rice cultivars. Japanese Journal of Breeding 3:33-43

Parnell F.R., Ayyangar G.N.R., Ramiah K., Ayyangar C.R.S. (1921) The inheritance of characters in rice, II. Memoirs of the Department of Agriculture in India, Botanical Series IX:185-208

Rao C.H., Seetharaman R. (1973) Genetic studies in pericarp and hull colour in rice. Indian J Genet Plant Breed 33:319-323

Ren F., Lu B.-R., Li S., Huang J., Zhu Y. (2003) A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. Theoretical and Applied Genetics 108:113-120

Saito A., Yano M., Kishimoto N., Nakagahra M., Yoshimura A., Saito K., Kuhara S., Ukai Y., Kawase M., Nagamine T., Yoshimura S., Ideta O., Ohsawa R., Hayano Y., Iwata N., Sugiura M. (1991) Linkage map of restriction fragment length polymorphism loci in rice. Japanese Journal of Breeding 41:665-670

Saitoh K., Onishi K., Mikami I., Thidar K., Sano Y. (2004) Allelic diversification at the *C* (*OsC1*) locus of wild and cultivated rice: Nucleotide changes associated with phenotypes. Genetics 168:997-1007

Sakamoto W., Ohmori T., Kageyama K., Miyazaki C., Saito A., Murata M., Noda K., Maekawa M. (2001) The *Purple leaf* (*Pl*) locus of rice: the *Plw* allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. Plant and Cell Physiology 42:982-991

Shobha Rani N., Pandey M.K., Prasad G.S.V., Kumar D. (2008) Inheritance of hull pigmentation and awning in short grain aromatic rice. Indian Journal of Crop Science 3:187-188

Shomura A., Izawa T., Ebana K., Ebitani T., Kanegae H., Konishi S., Yano M. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet 40:1023-1028

Sweeney M.T., Thomson M.J., Pfeil B.E., McCouch S. (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294

Sweeney M.T., Thomson M.J., Cho Y.G., Park Y.J., Williamson S.H., Bustamante C.D., McCouch S.R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genetics 3:e133

Tadege M., Sheldon C.C., Helliwell C.A., Upadhyaya N.M., Dennis E.S., Peacock W.J. (2003) Reciprocal control of flowering time by *OsSOC1* in transgenic Arabidopsis and by *FLC* in transgenic rice. Plant Biotechnology Journal 1:361-369

Takahashi M.E. (1957) Analysis of apiculus colour genes essential to anthocyanin colouration in rice. Journal of the Faculty of Agriculture, Hokkaido University 50:226-362

Takano-Kai N., Jiang H., Kubo T., Sweeney M., Matsumoto T., Kanamori H., Padhukasahasram B., Bustamante C., Yoshimura A., Doi K., McCouch S. (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics 182:1323-1334

Tan L., Li X., Liu F., Sun X., Li C., Zhu Z., Fu Y., Cai H., Wang X., Xie D., Sun C. (2008) Control of a key transition from prostrate to erect growth in rice domestication. Nature Genetics 40:1360-1364

Tripathi R.S., Rao M.J.B.K. (1979) Inheritance and linkage relationship of scent in rice. Euphytica 28:319-323

Wang E., Wang J., Zhu X., Hao W., Wang L., Li Q., Zhang L., He W., Lu B., Lin H., Ma H., Zhang G., He Z. (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40:1370-1374

Xiong L.Z., Liu K.D., Dai X.K., Xu C.G., Zhang Q. (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theoretical and Applied Genetics 98:243-251

Yu Y., Tang T., Qian Q., Wang Y., Yan M., Zeng D., Han B., Wu C.-I., Shi S., Li J. (2008) Independent losses of function in a polyphenol oxidase in rice: Differentiation in grain discoloration between subspecies and the role of positive selection under domestication. Plant Cell 20:2946-2959

Chapter 8:
The Genetic and Geographic Origin of the *badh2.1* FNP: Scenarios

Three possible scenarios will be presented that could explain the origin of the *badh2.1* FNP, all of which are consistent with the patterns of genetic variation found in modern *O. sativa*. The first scenario is that the mutation arose in the wild ancestor of cultivated rice, *O. rufipogon*. *O. rufipogon* is a complex species covering a large geographic range across Asia, with evidence of pre-domestication divergence into *Japonica*-like and *Indica*-like types (Second, 1982; Chen et al., 1993; Lu et al., 2002; Cheng et al., 2003; Hu et al., 2006; Londo et al., 2006; Caicedo et al., 2007; Duan et al., 2007; Rakshit et al., 2007). The *badh2.1* FNP may have arisen in a *Japonica*-like *O. rufipogon* ancestor and been selected by early humans during the domestication process. The overlapping ranges of the differentiated wild types, coupled with the fewer fertility barriers of early cultivars (Kovach et al., 2007; Sang and Ge, 2007; Vaughan et al., 2008) would have easily allowed the FNP to move into *Indica*-like domesticates in the region, resulting in the Jasmine types. Yet, this scenario is unlikely for several reasons. First, in our large survey of 192 *O. rufipogon* accessions from diverse regions across Asia, none contained the *badh2.1* allele. While the *badh2.1* allele has been identified in wild rice germplasm from Thailand (Vanavichit, 2004, 2007; Prathepha, 2009), it is most likely that these represent introgressions of the mutant allele into weedy rice populations growing in the vicinity of the large fields of Jasmine cultivars grown in Thailand (Bourgis et al., 2008). Also, the extreme reduction in diversity around the *badh2.1* allele compared to the wild type allele (Table 4.2; Figure 4.3) is evidence that the *badh2.1* FNP is a recent mutation having undergone a strong selective sweep. Finally, it is unlikely that fragrance would have

been selected early during the domestication of rice, since it does not appear to offer any selective advantage for agricultural production.

A second scenario for the origin of this mutation is that it arose post-domestication in an early *Japonica* cultivar. Early rice cultivars of both the *Indica* and *Japonica* varietal groups are thought to have had overlapping ranges across southern China and Southeast Asia. This "hybrid zone" would have provided an easy avenue for domestication-related alleles, as well as other important traits, to be transferred among early cultivars (Kovach et al., 2007; Sang and Ge, 2007; Vaughan et al., 2008). The *badh2.1* FNP may have arisen in southern China in an early *Japonica* cultivar, where it was then able to spread both south into the *Indica* cultivars of Southeast Asia via introgression and North along the Silk Road to the Kashmir region of India/Pakistan, where the *Group V* subpopulation is now endemic. Indeed, the prevalence of the *Japonica*-like *Group V* subpopulation in the Kashmir region, on a subcontinent that is otherwise largely *Indica*, suggests this genetic group arrived via westward-migrating human populations. The fragrance trait was likely highly valued in this region, where the environmental conditions favor optimal production of 2AP (Efferson, 1985; Itani et al., 2004), causing the *badh2.1* allele to become nearly fixed within the *Group V* subpopulation.

A final scenario is related to the second, except it involves a two-step process. First, the *badh2.1* FNP arose in an early *Japonica* cultivar and then migrated with humans along the Silk Road into the Kashmir region, where the *Group V* subpopulation was founded based on the fragrance trait. The *Group V* subpopulation would have flourished in relative geographic isolation from wild and other *O. sativa* populations, allowing it to accumulate rare alleles. The second step would then have been direct

introgression of the *badh2.1* allele into *Indica* varieties in Southeast Asia. This appears to be the most likely scenario, given that all tested *Indica* varieties carrying the *badh2.1* allele also carried polymorphisms unique to the *Group V* subpopulation within the introgressed region around *BADH2* (Figure 4.1B). It would also explain why the *tropical japonica* varieties, which are largely located in eastern Asia, lack these *Group V*-specific polymorphisms, since they would have inherited the *badh2.1* allele from the original undifferentiated *Japonica* ancestor instead of later from the *Group V* subpopulation.

Exactly when and where the introgression from *Group V* varieties into *Indica* occurred is beyond the scope of this study, but close examination of our data suggest the *badh2.1* allele in Jasmine-types was inherited from what are referred to as "traditional basmati" landraces, such as Taroari Basmati, HBC-19, and Karnal Local which were originally collected in the heart of the Kashmir region in Haryana, India (Aggarwal et al., 2002; Bhattacharjee et al., 2002; Nagaraju et al., 2002; Jain et al., 2006). This observation is based on a highly polymorphic SSR (TA repeat) in the 4th intron of the *BADH2* gene. Most *badh2.1*-bearing accessions of the *Group V* subpopulation have a unique allele at this SSR (11 repeats), while the "traditional basmati" landraces have a different allele (12 repeats). All tested *badh2.1*-bearing accessions of the *Indica* subpopulation carried the 12-repeat allele at this SSR locus (Supplemental Table 8.1), which may indicate the Jasmine-types inherited their fragrance allele from the traditional basmati-types that are to this day considered the highest quality basmati rice.

While in this study we could trace the origin of the fragrant allele in Jasmine varieties, we are unable to irrefutably determine which subpopulation within the *Japonica*

varietal group the *badh2.1* mutation originated. The extended haplotypes for the Fragrant *Japonica* and Fragrant *Group V* accessions are nearly identical to the ancestral Nonfragrant *Japonica* haplotype, with the exception of the FNP (Figure 4.1B). Yet, there are some polymorphisms that provide evidence that the *badh2.1* allele in the *tropical japonica* also arose in the *Group V* subpopulation. First, close examination of a highly variable SSR in the 4th intron of the gene indicates the fragrant *tropical japonica* accessions carry the SSR allele that is nearly fixed in the *Group V* accessions carrying the *badh2.1* allele (Supplemental Table 8.1). Also, the *tropical japonica* accessions with the *badh2.1* allele carry the "0" allele at extended haplotype position 69 (39kb downstream of *BADH2*; poly-T region with 8, 9, or 10 repeats) that is fixed within all *Group V* accessions with the *badh2.1* allele (Figure 4.2). This is confounded, though, by the fact that there are some nonfragrant *temperate* and *tropical japonica* accessions that also carry the "0" allele. Another hint suggesting a *Group V* origin is a 3 bp deletion found in the 5' UTR of most *tropical japonica* accessions carrying the *badh2.1* allele that is not found in any of the ancestral *Japonicas* nor *Group V* (Supplemental Figure 8.1). It is more likely that this allele arose more recently than the *badh2.1* FNP following introgression of *badh2.1* into *tropical japonica* from *Group V*. Finally, the EHH analysis for the *badh2.1* allele in *Group V*, *tropical japonica*, and *indica*, shows an elevated level of LD in *Group V* across the entire 5.3Mb region, while LD decays substantially across the region in *tropical japonica* and *indica*. The more rapid breakdown of LD in *tropical japonica* and *indica* suggest that the *badh2.1* allele was introgressed into these groups, with LD breakdown occurring over the course of historical recombination.

APPENDIX

Supplemental Table 8.1: Close-up of Key Polymorphisms in *BADH2* Gene

Accession	ID #	Subpop	Origin	IN1F_139	IN1F_587	IN1F_593	IN1R_315	IN2_88	IN2_164	IN2_362	IN2_374	IN4_62	IN4_386	IN4_398	IN4_451	IN5_388	IN5_390	IN5_398	IN6_41	IN6_193	IN6_288
Baber	9	*temperate japonica*	India	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Chinese	31	*temperate japonica*	China	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Early Wataribune	51	*temperate japonica*	Japan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Luk Takhar	103	*temperate japonica*	Afghanistan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Shinriki	143	*temperate japonica*	Japan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Tainan Iku 487	157	*temperate japonica*	Taiwan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Taipei 309	158	*temperate japonica*	Taiwan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Nipponbare	173	*temperate japonica*	Japan	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Randhanipagal (scented)	8183	*temperate japonica*	China	0	A	A	T	0	C	A	0	T	A	14	A	A	0	0	G	C	T
Chahora 144	27	*tropical japonica*	Pakistan	0	A	A	T	2	C	A	0	T	T	2	A	A	0	0	G	C	T
IRAT 177	73	*tropical japonica*	French Guiana	0	A	A	T	2	C	A	0	T	T	2	A	A	0	0	G	C	T
ARC 6070	8177	*Group V*	India	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
ARC 13829	F58	*Group V*	India	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
ARC 13523	8176	*Group V*	India	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Farangey	8181	*Group V*	Bhutan	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Pato De Gallinazo	128	*temperate japonica*	Australia	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Trembese	165	*tropical japonica*	Indonesia	0	A	A	T	2	C	A	0	N	T	4	A	A	0	0	G	C	T
DA 5	38	*tropical japonica*	Bangladesh	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Kotobuki Mochi	95	*tropical japonica*	Japan	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Honduras	65	*tropical japonica*	Honduras	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
OS6	120	*tropical japonica*	Nigeria	0	A	A	T	2	C	A	0	N	T	4	A	A	0	0	G	C	T
Ble Ma Mua	8395	*tropical japonica*	Vietnam	0	A	A	T	2	C	A	0	T	T	4	A	A	0	0	G	C	T
Chhote Dhan	F79	*Group V*	Nepal	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Basmati 150	2239	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Tchampa	160	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Basmati 370B	2237	*Group V?*	Pakistan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
NORIN 20	113	*temperate japonica*	Japan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Hu Lo Tao	67	*temperate japonica*	China	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
GEUMOBYEO	56	*temperate japonica*	Korea	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Kamenoo	83	*temperate japonica*	Japan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Jouiku 393G	79	*temperate japonica*	Japan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Koshihikari	94	*temperate japonica*	Japan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Ta Mao Tsao	155	*temperate japonica*	China	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Ostiglia	121	*temperate japonica*	Argentina	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Darmali	41	*temperate japonica*	Nepal	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Gotak Gatik	60	*temperate japonica*	Indonesia	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
NPE 835	115	*temperate japonica*	Pakistan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
TA HUNG KU	154	*temperate japonica*	China	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Leung Pratew	102	*temperate japonica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Agostano	1	*temperate japonica*	Italy	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Oro	118	*temperate japonica*	Chile	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Romeo	134	*temperate japonica*	Italy	0	A	A	N	2	C	A	0	T	T	6	A	A	0	0	G	C	T
KU115	96	*tropical japonica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Miriti	107	*tropical japonica*	Bangladesh	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
NPE 844	116	*tropical japonica*	Pakistan	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Kinastano	92	*tropical japonica*	Philippines	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Moroberekan	108	*tropical japonica*	Guinea	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Dourado Agulha	46	*tropical japonica*	Brazil	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
IAC 25	69	*tropical japonica*	Brazil	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
LAC 23	99	*tropical japonica*	Liberia	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
RT 1031-69	135	*tropical japonica*	Zaire	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Canella De Ferro	23	*tropical japonica*	Brazil	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Cuba 65	37	*tropical japonica*	Cuba	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
L-202	98	*tropical japonica*	USA	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Davao	42	*tropical japonica*	Philippines	0	A	A	T	2	C	A	0	T	T	6	A	A	0	0	G	C	T
Cs-M3	36	*temperate japonica*	USA	0	A	A	T	2	C	A	0	T	T	8	A	A	0	0	G	C	T
JC1	F3	*Group V*	India	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
BASFUL 714	8160	*Group V*	Bangladesh	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
Jasmine Scented	8516	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
Somaly Krar-oob	8521	*indica*	Cambodia	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
HNN	8155	*indica*	Laos	0	A	A	N	2	C	A	0	T	T	4	A	T	5	3	G	C	T
Chao do	8147	*indica*	Laos	0	A	A	N	2	C	A	0	T	T	4	A	T	5	3	G	C	T
WC 6	169	*temperate japonica*	China	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
Keriting Tingii	87	*tropical japonica*	Indonesia	0	A	A	T	2	C	A	0	T	T	4	A	T	5	3	G	C	T
Kalijira	8503	*Group V*	Bangladesh	0	A	A	N	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Sadri Ghermes	8514	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Blomberg	2254	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Chini sakkar	2258	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
TARAORI	2278	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Pusa basmati-1	2279	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
HBC-19	2280	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
HARANDI 379	8174	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Basmati 802	8232	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
HKR93(401)	2277	*Group V*	India	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Nang Thom Cho Dao	8127	*indica*	Vietnam	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
MA WAINE OHN	8189	*indica*	Myanmar	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
PAWSANHMWE	8191	*indica*	Myanmar	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Yun Aromatic Glutinous	8222	*indica*	China	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
RTS12	136	*indica*	Vietnam	N	N	N	N	2	C	A	0	T	T	6	A	T	5	3	N	N	N
GUOR LAO	8187	*indica*	Vietnam	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Q74	8126	*indica*	Malaysia	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Champa tong 54	F21	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
PTT1	8131	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Rathuwee	132	*indica*	Sri Lanka	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
BOKEHMWE	8186	*indica*	Myanmar	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Khao Dawk Mali 105	4878	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
JASMINE 85	8224	*indica*	USA	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
TAINUNG SEN 20	8193	*indica*	Taiwan	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Nho Thom	8141	*indica*	Vietnam	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
RD33	8130	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
RD6	8132	*indica*	Thailand	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
OM3536	8135	*indica*	Vietnam	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T

Supplemental Table 8.1 (continued)

P. Gadis	J55	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
P. Pulut Mbau	J176	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Bidor	8362	*tropical japonica*	Malaysia	3	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
Coreng	8364	*tropical japonica*	Malaysia	3	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
P. Ciu	J186	*tropical japonica*	Indonesia	0	A	A	T	2	C	A	0	T	T	6	A	T	5	3	G	C	T
JC220	8188	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Ambemohar 1	2221	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
AMBEMOHOR 157	8157	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Gogo Lempuk	59	*Group V*	Indonesia	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Gerdeh	55	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati bahar	2226	*Group V*	India	0	A	A	N	2	C	A	0	T	T	8	A	T	5	3	G	C	T
ARC 6011	8158	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 6113	F61	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
SADRI	8175	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
ARC 10352	5	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Dulhemiya	2262	*Group V*	India	0	A	A	N	2	C	A	0	T	T	8	A	T	5	3	G	C	T
DOMSIAH	8171	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
DA13	8169	*Group V*	Bangladesh	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Pankhari 203	F38	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
DOM-SOFID	45	*Group V*	Iran	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 1	F14	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 217	14	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 213	F16	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 370B	F18	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati sathi	F33	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Kitrana 508	93	*Group V*	Madagascar	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
N12	112	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BPT1235	2248	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Barhai 1	2253	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Chini guri	2259	*Group V*	Bangladesh	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
D66	2264	*Group V*	Bangladesh	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BASMATI 6311	8162	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BASMATI C 622	8163	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Aus Basmati	8195	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Chahora 144	F20	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 5836	F63	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BASMATI 370	8185	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati Lal	8201	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BARA 413	8159	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Sathi Basmati	8231	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 2000	8136	*Group V*	Pakistan	0	A	A	T	N	N	N	N	T	T	8	A	T	5	3	G	C	T
Basmati 372A	F19	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
HKR 228	2271	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 122	F75	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Karnal Basmati	8208	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 410	8200	*Group V*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Basmati 802	8206	*Group V*	Pakistan	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Della	391	*indica*	USA	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
BC5-55	8230	*indica*	India	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Beonjo	15	*temperate japonica*	Korea	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
P. Atok	J152	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
P. Ikeng	J187	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
P. Telengusan	J178	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
P. Pulut Longbanga	J13	*tropical japonica*	Indonesia	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Azucena	7737	*tropical japonica*	Philippines	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Asse Y Pung	8	*tropical japonica*	Philippines	3	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Arias	7	*tropical japonica*	Indonesia	0	A	A	T	2	C	A	0	T	T	8	A	T	5	3	G	C	T
Kun-Min-Tsieh-Hunan	97	*indica*	China	0	C	A	C	0	T	T	3	A	A	4	G	A	0	0	A	G	C
BYAKKOKU Y 5006 SELN	21	*indica*	Australia	0	C	A	C	0	T	T	3	A	A	4	G	A	0	0	A	G	C
Hsia-Chioh-Keh-Tu	66	*indica*	Taiwan	0	C	A	C	0	T	T	3	A	A	6	G	A	0	0	A	G	C
Shuang-Chiang	146	*indica*	Taiwan	0	A	A	T	0	T	T	3	A	A	6	G	A	0	0	A	G	C
EH-1A-Chiu	52	*temperate japonica*	Taiwan	0	N	N	C	0	T	T	3	A	A	6	G	A	0	0	A	G	C
Basmati	12	*Group V*	Pakistan	0	C	A	C	0	T	T	3	A	A	6	G	A	0	0	A	G	C
Chiem Chanh	30	*indica*	Vietnam	0	C	A	C	0	T	T	3	A	A	6	G	A	0	0	A	G	C
Taducan	163	*indica*	Philippines	N	C	C	C	0	T	T	3	A	A	6	G	A	0	0	A	G	C
Basmati mehtrah	F80	*Group V*	India	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Basmati Mehtrah	2225	*Group V*	India	0	C	C	N	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Ghati Kamma Nangarhar	58	*aus*	Afghanistan	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Jhona 349	78	*aus*	India	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
BEGUMI 302	8178	*Group V*	Pakistan	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
JHONA 349	AUS2	*aus*	India	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Baugamon 14	11	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Basmati 1	13	*aus*	Pakistan	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
DUMAI	AUS7	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
GERDEH	AUS8	*aus*	Iran	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
LAKHI PURI	AUS9	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Firooz	53	*Group V*	Iran	0	C	N	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Bico Branco	16	*Group V*	Brazil	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
DZ78	50	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Kasalath	85	*aus*	India	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Champa Tong 54	28	*aus*	Thailand	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Wells	170	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
Basmati 334	F17	*Group V*	Pakistan	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
DA16	AUS1	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
KALUBALA VEE	AUS4	*aus*	Sri Lanka	0	C	C	C	0	T	T	3	A	A	14	G	A	0	0	A	G	C
MAHIN KUNDUZ	8182	*Group V*	Afghanistan	0	C	C	C	0	T	T	3	A	A	15	G	A	0	0	A	G	C
TKM6	162	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
CO 25	34	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
CO18	35	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
RTS4	138	*indica*	Vietnam	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Pappaku	126	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Guan-Yin-Tsan	61	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
JC149	77	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C

Supplemental Table 8.1 (continued)

Chau	29	*indica*	Vietnam	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Zhe 733	171	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Kiang-Chou-Chiu	90	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
O-LUEN-CHEUNG	117	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Dee Geo Woo Gen	43	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Peh-Kuh	129	*indica*	Taiwan	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Peh-Kuh-Tsao-Tu	130	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Taichung Native 1	156	*indica*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Teqing	161	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Gharib	57	*indica*	Iran	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Jaya	76	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Pratao	F4	*indica*	Brazil	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Oryzica Llanos 5	119	*indica*	Colombia	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Basmati 213	2232	*indica*	Pakistan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Seratoes Hari	141	*indica*	Indonesia	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Jambu	75	*tropical japonica*	Indonesia	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Binulawan	17	*indica*	Philippines	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
IR 8	72	*indica*	Philippines	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
IR64	5133	*indica*	Philippines	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Pao-Tou-Hung	125	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Zhenshan 2	172	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Padi Kasalle	122	*tropical japonica*	Indonesia	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Ai-Chiao-Hong	3	*indica*	China	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Tondok	164	*tropical japonica*	Indonesia	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Dholi Boro	4984	*indica*	Bangladesh	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Mudgo	110	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Sintane Diofor	148	*indica*	Burkino Faso	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Khao Gaew	88	*aus*	Thailand	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
DV85	49	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
DA8	AUS3	*aus*	Bangladesh	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
JC148	AUS5	*aus*	India	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
BAILAM	AUS6	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
ARC 7229	6	*aus*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
BJ 1	18	*aus*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
FR13 A	4891	*aus*	India	N	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Dhala Shaitta	44	*aus*	Bangladesh	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Phudugey	131	*aus*	Bhutan	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
T 1	152	*aus*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
IR 36	71	*indica*	Philippines	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Chuan 4	33	*aus*	Taiwan	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
BLACK GORA	19	*aus*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Mehr	105	*aus*	Iran	0	N	N	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Lemont	101	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
S4542A3-49B-2B12	139	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Sinampaga Selection	147	*tropical japonica*	Philippines	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
SINAGUING	149	*tropical japonica*	Philippines	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
B6616A4-22-Bk-5-4	167	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	C	C
ARC 10177	4	*aus*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
RTS14	137	*indica*	Vietnam	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Khao Hawm	89	*tropical japonica*	Thailand	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
N 22	111	*tropical japonica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Fortuna	54	*tropical japonica*	USA	0	C	C	N	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Saturn	140	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Sultani	150	*tropical japonica*	Egypt	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Bogi Joha	2250	*indica*	India	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Nova	114	*tropical japonica*	USA	0	C	C	C	0	T	T	3	A	A	16	G	A	0	0	A	G	C
Kalamkati	81	*aus*	India	0	C	C	C	0	T	T	3	A	A	18	G	A	0	0	A	G	C

This table zooms in on the *BADH2* gene in the panel of 252 *O. sativa* accessions used in the Kovach et al., 2009 study. A subset of the AIPs are shown, along with key polymorphisms for determining the origin of the *badh2.1* allele. The accessions outlined in a red box are traditional basmati landraces (or have a traditional landrace in their pedigree), and are believed to be the 'original' or 'oldest' basmati varieties. Notice that the majority of Jasmine accessions (highlighted in light blue) share the same allele at the highly polymorphic SSR as these traditional basmati accessions. Also notice that the majority of *tropical japonica* accessions with the *badh2.1* allele have the same number of repeats at the SSR as the majority of contemporary *Group V* accessions. Finally, a 3 bp deletion was detected in the majority *tropical japonica* accessions possessing *badh2.1*, while this deletion is absent from *Group V* and non-fragrant *tropical japonica*.

REFERENCES

Aggarwal, Aggarwal R., Shenoy, Shenoy V., Ramadevi, Ramadevi J., Rajkumar, Rajkumar R., Singh, Singh L. (2002) Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent-AFLP. Theoretical and Applied Genetics 105:680-690

Bhattacharjee P., Singhal R.S., Kulkarni P.R. (2002) Basmati rice: a review. International Journal of Food Science & Technology 37:1-12

Bourgis F., Guyot R., Gherbi H., Tailliez E., Amabile I., Salse J., Lorieux M., Delseny M., Ghesquière A. (2008) Characterization of the major fragance gene from an aromatic *japonica* rice and analysis of its diversity in Asian cultivated rice. Theoretical and Applied Genetics 117:353-368

Caicedo A.L., Williamson S.H., Hernandez R.D., Boyko A., Fledel-Alon A., York T.L., Polato N.R., Olsen K.M., Nielsen R., McCouch S.R., Bustamante C.D., Purugganan M.D. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genetics 3:e163

Chen W.B., Nakamura I., Sato Y.I., Nakai H. (1993) Distribution of deletion type in cpDNA of cultivated and wild rice. Japanese Journal of Genetics 68:597-603

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: Based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67-75

Duan S., Lu B., Li Z., Tong J., Kong J., Yao W., Li S., Zhu Y. (2007) Phylogenetic analysis of AA-genome *Oryza* species (*Poaceae*) based on chloroplast, mitochondrial, and nuclear DNA sequences. Biochemical Genetics 45:113-129

Efferson J.N.E. (1985) Rice quality in world markets. In: IRRI (ed) Rice Grain Quality and Marketing. International Rice Research Institute, Manila, Philippines, pp 1-13

Hu H., Mu J., Zhang H.-J., Tao Y.-Z., Han B. (2006) Differentiation of a miniature inverted transposable element (MITE) system in Asian rice cultivars and its inference for a diphyletic origin of two subspecies of Asian cultivated rice. Journal of Integrative Plant Biology 48:260-267

Itani T., Tamaki M., Hayata Y., Fushimi T., Hashizume K. (2004) Variation of 2-acetyl-1-pyrroline concentration in aromatic rice grains collected in the same region in Japan and factors affecting its concentration. Plant Production Science 7:178-183

Jain N., Jain S., Saini N., Jain R. (2006) SSR analysis of chromosome 8 regions associated with aroma and cooked kernel elongation in Basmati rice. Euphytica 152:259-273

Kovach M.J., Sweeney M.T., McCouch S.R. (2007) New insights into the history of rice domestication. Trends in Genetics 23:578-587

Londo J.P., Chiang Y.-C., Hung K.-H., Chiang T.-Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proceedings of the National Academy of Sciences 103:9578-9583

Lu, Lu B.R., Zheng, Zheng K., Qian, Qian H., Zhuang, Zhuang J. (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics 106:101-106

Nagaraju J., Kathirvel M., Kumar R.R., Siddiq E.A., Hasnain S.E. (2002) Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by

using fluorescence-based ISSR-PCR and SSR markers. Proceedings of the National Academy of Sciences of the United States of America 99:5836-5841

Prathepha P. (2009) The fragrance (*fgr*) gene in natural populations of wild rice (*Oryza rufipogon* Griff.). Genetic Resources and Crop Evolution 56:13-18

Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N., Terauchi R. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theoretical and Applied Genetics 114:731-743

Sang T., Ge S. (2007) The puzzle of rice domestication. Journal of Integrative Plant Biology 49:760-768

Second G. (1982) Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Japanese Journal of Genetics 57:25-57

Vanavichit A. (2004) Discovering genes for rice grain aroma. Proceedings of the 1st International Conference on Rice for the Future, Bangkok, Thailand

Vanavichit A. (2007) Molecular diversity and evolution of aromatic rice in Thailand. International Training Workshop, The Conservation and Utilization of Tropical/Subtropical Plant Genetic Resources. Taiwan Agricultural Research Institute, Council of Agriculture (TARIC), Taiwan, pp 131-134

Vaughan D.A., Lu B.-R., Tomooka N. (2008) The evolving story of rice evolution. Plant Science 174:394-408

Chapter 9:
Characterization of RC-Mediated Regulation of Proanthocyanidin Biosynthesis as a Prerequisite for a Novel Transgene Containment Strategy in Rice

INTRODUCTION

Red rice: Noxious weed or valuable reservoir of genetic variation?

Rice (*Oryza sativa*) is a staple food for a huge portion of the human population. With the most rapid world population growth occurring in regions where rice is the major source of calories, the global demand for rice is expected to increase markedly over the next several decades (Maclean et al., 2002). The United States produced over ten million tons of rice on nearly three million acres in 2008 and is poised to take advantage of the global need for rice through its export markets, which already account for almost 1.4 billion dollars in annual export revenues (FAO, 2009). In addition, there has been a steady increase in rice consumption in the U.S., causing domestic and residual use projections to reach record levels for 2010/2011 (USDA, 2009).

One of the most economically significant pest problems facing the U.S. rice industry is weedy red rice. Red rice causes losses of over $50 million per year (Gealy et al., 2002), largely due to a lower market value for grain lots that are contaminated with red seed. While red rice grains look similar to cultivated forms of rice, these grains cook differently and the pigmented bran layer is difficult to remove, requiring longer milling times and resulting in more broken grains (Kwon et al., 1992). In addition to red seed color, weedy rice biotypes exhibit dormancy, shattering, and are able to interbreed freely with cultivated, white-grained rice, causing extensive control

difficulties and allowing viable seeds to persist as dormant propagules in the field for many years (Goss and Brown, 1939). Further, red rice plants are close mimics of cultivated varieties in the field and exist in close proximity to cultivated populations, making it nearly impossible to utilize herbicides that effectively target the weeds without damaging the crop (Burgos et al., 2003). Weedy red rice species emerge faster, grow taller, and produce more tillers than cultivated varieties, causing competition for light and nutrients (Diarra et al., 1985; Burgos et al., 2000). These production constraints, along with the decreased market value of rice contaminated with red grains, results in substantial economic losses for growers. Red rice is one of the top ten most troublesome weeds and the most competitive grass weed in rice production (Webster, 2000), which has led to the classification of red rice as a noxious weed in the United States.

Despite the negative aspects of red rice, wild ancestors of cultivated rice, such as *O. rufipogon* and *O. nivara* represent valuable reservoirs of genetic variation with the ability to fuel rice improvements far into the future (Tanksley and McCouch, 1997; Zamir, 2001; McCouch, 2004). In some parts of the world, red-grained cultivars are preferred for their taste, texture, and religious or medicinal value. Red and black varieties contain up to 37% more protein, 17% more crude fiber, and are richer in lysine and vitamin B_1 than conventional varieties (Duffy, 2001). The flavonoid pigments responsible for grain coloration have potential benefits for human nutrition, as they have been shown to act as powerful antioxidants, reducing atherosclerotic plaque formation and decreasing the risk of heart disease (Ling et al., 2001). From an agronomic perspective, the study of proanthocyanidin synthesis could allow for manipulation of proanthocyanidin content in forage crops, as these compounds have been found to reduce bloat in grazing ruminants (MacKown et al., 2008).

Red Rice Weed Control

Traditional paddy rice cultivation, in which seedlings are transplanted into flooded soil, not only provides irrigation for the developing rice plants, but the standing water is also a very effective means of weed control. However, in the U.S., Africa, and in many parts of Asia, rice producers use primarily direct-seeding methods. The recent move away from the paddy system in regions such as Malaysia, Thailand, and the Philippines is associated with diminishing supplies of fresh water and increasing costs associated with the labor-intensive seedling preparation and transplanting activities (Serrano, 1975). The change to direct-seeding is almost always followed by a dramatic increase in weed difficulties, due to reduced manual inspection of plants going into the field, reduced hand-weeding, and minimal row management (Moody, 1992). A concomitant increase in herbicide use generally follows the adoption of direct-seeding, despite increasing herbicide prices (Erguiza et al., 1990).

In the U.S., where direct-seeding methods are used exclusively, rice producers rely on a variety of herbicides for effective weed control. The development of Clearfield® rice varieties, which contain a mutation conferring resistance to the imidazolinone herbicide Newpath®, has helped to diversify current weed control strategies. Clearfield® plants are highly resistant to the Newpath® herbicide, while red rice weeds are highly susceptible (Sanders et al., 1998). However, a multi-year study at the Rice Research and Extension Center in Stuttgart, Arkansas revealed that the Clearfield® varieties were in fact transmitting their herbicide-resistance trait to nearby red rice populations (Burgos et al., 2003) suggesting that without some method of inhibiting gene flow, the effectiveness of Clearfield® varieties will be short-lived. Transgenic herbicide-resistant varieties offer another option for ameliorating weed problems in our herbicide-dependent rice production industry. The risk of transferring herbicide-

resistant transgenes to wild or weedy relatives is similar to the risk of transmitting the mutant gene found in Clearfield® varieties, but the public's low tolerance of the risk associated with transgenics has thus far impeded the commercial release of transgenic rice varieties.

Gene flow of herbicide-resistant pollen from rice cultivars to rice weeds or wild relatives poses two primary risks. The first is the creation of weedy biotypes with acquired herbicide-resistance. Although studies examining gene flow between wild or weedy rice and cultivated rice usually report less than 1% outcrossing rates (Coffman and Herrera, 1980), the intense selection pressure that herbicide application exerts on the red rice population would increase the chances of an herbicide-resistant weed becoming prevalent (Gealy et al., 2003). The second risk is the threat of contaminating the genetic integrity of wild *O. sativa* ancestors, such as *O. rufipogon* and *O. nivara*, if transgenics are grown in the center of diversity for *Oryza*. Wild ancestors of our modern crop varieties contain a vast wealth of hidden genetic variation, which can be used by plant breeders to improve productivity, stress-resistance, and other key traits to crop production and consumption (Tanksley and McCouch, 1997; McCouch, 2004). The importance of this pool of genetic diversity has long been recognized as essential to the future of crop improvement (Bessey, 1906). Compromising the wild germplasm of rice would deprive future generations of these valuable sources of genetic diversity that will be necessary to feed an ever-growing human population. Thus, for transgenic herbicide-resistant rice varieties to be used in an economically and environmentally responsible way, a method is needed to block transmission of transgenes to weedy "red rice" and wild rice relatives.

The cloning and characterization of *Rc*

Red pericarp is characteristic of all wild *Oryza* species and many of the early landrace varieties. This feature of the grain readily differentiates ancestral rice genotypes from modern, high-yielding varieties and offers a visible marker that easily distinguishes between these pools of germplasm. The pigments in red rice pericarp are proanthocyanidins, also called condensed tannins (Oki et al., 2002). Proanthocyanidins, which are polymers of flavan-3,4-diols and flavan-3-ols, are derived from the general flavonoid pathway in plants. These and related compounds are responsible for a variety of biological functions, including pest and pathogen defense and protection against UV radiation (Winkel-Shirley, 2001).

There are two loci required for red pericarp pigmentation, identified by classical genetic analysis, *Rd* and *Rc* (Kato & Ishikawa, 1921; Kinoshita, 1998). The *Rc* gene was recently cloned found to encode a Myc-type transcription factor containing a basic helix-loop-helix (bHLH) domain (Sweeney et al., 2006). When the dominant, functional forms of *Rc* and *Rd* are present, rice pericarp accumulates proanthocyanidins, giving the grains a red color. Plants that have a nonfunctional *rd*, but a functional *Rc* have brown pericarp, while plants with the mutant allele *rc*, regardless of the state of *Rd*, do not accumulate proanthocyanidins and consequently have white grains.

Sweeney et al. (2006) determined this phenotypic difference was caused by a 14 base pair deletion in the coding region of the *Rc* gene, causing a frame-shift and truncation of the resulting protein before the bHLH domain. *Rc* expression was limited to panicles and seeds, and was not detected in vegetative tissues. A further survey of rice germplasm has revealed that over 97% of white-grained genotypes, regardless of

subpopulation identity, carry this 14-bp deletion (M. Sweeney, Cornell University, unpublished data). An additional mutation within this gene (creating the *Rc-s* allele) is responsible for the remaining 3% of white genotypes. While most of the plants with the *Rc-s* allele produced white seeds, several genotypes with this allele also produce seeds with light red pericarp (Sweeney et al., 2006). These studies indicate that a mutation in a single transcription factor can alter the pericarp color of a cereal grain, with far-reaching evolutionary consequences.

Transcriptional regulation of the flavonoid pathway in plants: An unfinished story

Virtually all the biosynthetic enzymes of the flavonoid pathway have been identified and their corresponding genes have been cloned from several species, including rice (Atkins et al., 1991; Holton and Cornish, 1995; Reddy et al., 1996; Winkel-Shirley, 2001; Druka et al., 2003). In addition to the biosynthetic genes that catalyze the synthesis of flavonoids, a number of regulatory elements that attenuate the expression of these biosynthetic genes, like *Rc*, have also been identified in multiple species (reviewed in (Holton and Cornish, 1995)).

The regulation of flavonoid accumulation is a complex and intricate process, and studies of this pathway have contributed greatly to our understanding of the regulation of secondary metabolism in plants. Flavonoids are particularly interesting features of secondary metabolism, due to their roles in the production of pigments and for providing a variety of useful compounds to human health. Since plant organ pigmentation is an easily observable phenotype, the flavonoid pathway has been an ideal system for the study of metabolic pathways and gene regulation. The extensive genetic and molecular characterization of the biosynthetic and regulatory genes of this pathway has provided some of the most detailed knowledge of gene interactions in

plants (Tuerck and Fromm, 1994). This knowledge has led to practical application through metabolic engineering of favorable traits into crop systems, such as tomato and alfalfa ((Bovy et al., 2002; Ray et al., 2003).

Despite the limited successes in metabolic engineering of flavonoid traits, there remain significant gaps in knowledge regarding the exact genetic and molecular mechanisms regulating the biosynthesis of molecules in this pathway. While many of the biosynthetic genes are highly conserved across plant species, the temporal and spatial expression of these genes varies greatly both within and between species (Quattrocchio et al., 1993; Holton and Cornish, 1995). The organ, tissue, and cell-specific accumulation of flavonoids in plants has been attributed to both the combinatorial interactions between regulatory elements and to variation in the promoters of the biosynthetic genes, on which the regulatory elements act.

Combinatorial interactions

In most cases, the accumulation of anthocyanins or proanthocyanidins in any plant organ requires the presence of at least two families of regulatory proteins: Myc-type proteins containing a basic helix-loop-helix (bHLH) domain, and R2R3 Myb proteins. Genes of these two families are present in numerous copies in plant genomes and encode similar proteins with varying expression patterns. Studies in maize revealed a member from each of these two families of transcriptional regulators was necessary to activate anthocyanin biosynthesis in all tissues and a direct physical interaction between these two proteins was required for activation (Goff et al., 1992). Subsequent research has revealed significant variations on this general theme, and indicates the need for further analysis of flavonoid pathway transcriptional regulation.

First, ectopic expression of *bHLH* and *Myb* genes causes transcriptional activation of flavonoid synthesis in some, but not all tissues. Expression of the *Lc* (maize; bHLH) and *C1* (maize; Myb) genes in *Arabidopsis* induces anthocyanin production in all organs, while in tobacco (*Nicotiana tabacum*), only the floral organs become pigmented (Lloyd et al., 1992). Similarly, though expression of *Lc* causes pigmentation of all organs in *Arabidopsis*, the *Lc* orthologs from *Antirrhinum* (*Delila*) and *Perilla* (*Myc-RP*) are unable to induce any anthocyanin synthesis in *Arabidopsis* despite high sequence similarity between these proteins (Gong et al., 1999). While *Lc* alone causes anthocyanin production in both vegetative and reproductive tissues of tomato, only reproductive tissues of tobacco are affected by *Lc* expression (Mooney et al., 1995). Also, when the petunia *bHLH* homolog, *Jaf13*, was introduced into maize *bHLH* mutants, it was unable to activate anthocyanin synthesis (Quattrocchio et al., 1998). These results suggest that interactions between bHLH proteins and their partner Myb proteins are very specific and the expression of these genes is restricted to certain tissues and developmental stages, which may vary by species. It also highlights the importance of the combinatorial interactions between these two families of transcription factors needed to impart transcriptional activation of the flavonoid pathway.

In addition to the variability outlined above, there is also a major discrepancy between *Rc* and its ortholog in maize, *Intensifier1* (*IN1*). While IN1 is a negative regulator of the flavonoid pathway (Burr et al., 1996), RC, like TT8 (*Arabidopsis*), and AN1 (petunia), is a positive regulator of proanthocyanidin biosynthesis (Nesi et al., 2000; Spelt et al., 2000; Sweeney et al., 2006).

The combinatorial interactions of bHLH and Myb proteins for regulation of the flavonoid pathway may only be part of a larger picture. Work in maize, snapdragon, petunia, and *Arabidopsis* has revealed that bHLH and Myb proteins are only two of potentially several factors involved in the control of flavonoid biosynthetic genes. Additional molecular factors implicated in bHLH family-mediated flavonoid biosynthetic gene regulation include WD40 proteins, such as the *TTG1* gene from *Arabidopsis*, *AN11* gene from petunia, and *PAC1* gene from maize (de Vetten et al., 1997; Bharti and Khurana, 2003; Carey et al., 2004). While TTG1 has been shown to physically interact with a bHLH protein in *Arabidopsis* (Sompornpailin et al., 2002), AN11 from petunia was found to be localized in the cytosol (de Vetten et al., 1997). In petunia and *Arabidopsis*, other regulatory factors, such as TT1, a zinc finger protein (Sagasser et al., 2002), and AN2, a homeodomain protein (Kubo et al., 1999), have also been described where loss of function results in a complete lack of pigmentation. This further demonstrates that both more general and more specific understandings of bHLH-mediated transcriptional activation in diverse plant species are needed if we are to constructively harness the flavonoid and other metabolic pathways for applications in agriculture and medicine.

DNA binding

A second gap in our knowledge has to do with the ability of bHLH proteins to bind DNA and the importance of this binding for transcriptional activation. Studies in mammals have shown that bHLH proteins are capable of binding to DNA *cis*-elements called "E-boxes" (CANNTG) in the promoters of their target genes (Ellenberger et al., 1994; Brownlie et al., 1997; Atchley et al., 1999), and that this binding site is conserved for plant bHLH proteins (sometimes referred to as "G-boxes") (Kawagoe and Murai, 1992; de Pater et al., 1997; Loulergue et al., 1998). Yet, while some bHLH

proteins are able to bind DNA, others lack this ability but are still able to activate the flavonoid pathway. The *Delila* (snapdragon), *Myc-RP* (*Perilla*), *RAP-1* (*Arabidopsis*), as well as many other bHLH genes have been shown to bind DNA (de Pater et al., 1997; Gong et al., 1999). In contrast, *Lc* (maize) does not bind DNA (Hernandez et al., 2004), yet is fully functional as a flavonoid pathway transcriptional activator. These proteins all share highly homologous DNA-binding regions, demonstrating again that sequence similarity is not sufficient to designate protein function.

The variability in the capacity of bHLH proteins to bind DNA directly may be a function of their interactions with Myb proteins and other molecular factors. Myb proteins contain three α-helices, forming a helix-turn-helix structure, with the third helix making sequence-specific contacts with DNA (Sainz et al., 1997; Grotewold et al., 2000). The *P* and *C1* genes from maize encode Myb proteins that bind directly to CC(T/A)ACC sites in the *A1* promoter for activation of the *Al* biosynthetic gene (which encodes the maize DFR) (Grotewold et al., 1994). The C1 Myb protein can also bind to a variety of sequences that resemble a consensus site A(C/A)C(T/A)A(C/A)C (Sainz et al., 1997). Other consensus sites for Myb protein-binding have been experimentally described in the promoters of multiple flavonoid biosynthetic genes(Tuerck and Fromm, 1994; Lesnick and Chandler, 1998). It is therefore likely that DNA binding by the bHLH, Myb, or both transcription factors together may dictate the tissue-specific activation of their target biosynthetic genes.

In summary, the biosynthesis of flavonoids in plants appears to be regulated by a complex suite of molecular factors. Interactions between the transcription factors and between these factors and their target DNA promoters dictate the exact tissue-specificity of flavonoid pathway regulation. While studies in other systems have

provided us with a working model, the variation in the regulation of the flavonoid metabolic pathway between species necessitates the characterization of this pathway in rice if this pathway is to be manipulated for rice improvement.

Toward a new strategy for transgene containment in rice

In addition to generating new, basic knowledge about transcriptional regulation in plants gained from the research presented here will also lay the foundation for developing a novel method of weed control in rice. This research will contribute to the initial stage of development of a form of herbicide-resistant rice that precludes the transfer of transgenic herbicide-resistance to weedy red rice genotypes. Because all modern rice varieties have white pericarp due to a nonfunctional rc transcription factor, while all wild and weedy rice examined to date has red pericarp due to a functional RC protein, there exists a pragmatic opportunity to develop a novel method of transgene containment. This strategy involves the following logic. A gene under the control of an RC-responsive promoter would be inactive in white (*rc*) cultivated rice pericarp tissue (just as the proanthocyanidin biosynthetic genes are present but inactive in the pericarp of these cultivars). However, this same transgene construct in a wild or weedy rice genotype (*Rc*) would be activated by the functional RC protein and would be expressed. The proposed research will therefore allow for the engineering of a transgenic construct containing both an herbicide-resistance gene under the control of the appropriate promoter and an embryo-lethal gene under the control of the RC-responsive promoter. This embryo-lethal gene can encode either a biological toxin or an RNAi construct targeting a gene necessary for embryo/seed development and survival. This mechanism would ensure that if transgenic pollen with this construct fertilizes any panicles from rice genotypes with a functional RC, the grains from these genotypes would be rendered sterile and unable to produce viable plants. This sterility

would be specific to red rice, while white cultivated rice would be fertile and could be saved and replanted by farmers and seed producers. The "contained" herbicide-resistance provided by this construct would not only protect against the risk of creating rice "superweeds", but would also allow for the preservation of precious wild germplasm for future breeding efforts. The diversity of flavonoid pathway transcriptional regulation in the various plant species studied to date highlights the need for precise characterization of this system in rice, in order to reliably develop the "contained" transgene containment concept.

OBJECTIVES

The objectives of this study are to test the following model of proanthocyanidin biosynthesis regulation by RC in the rice pericarp (Figure 9.1):

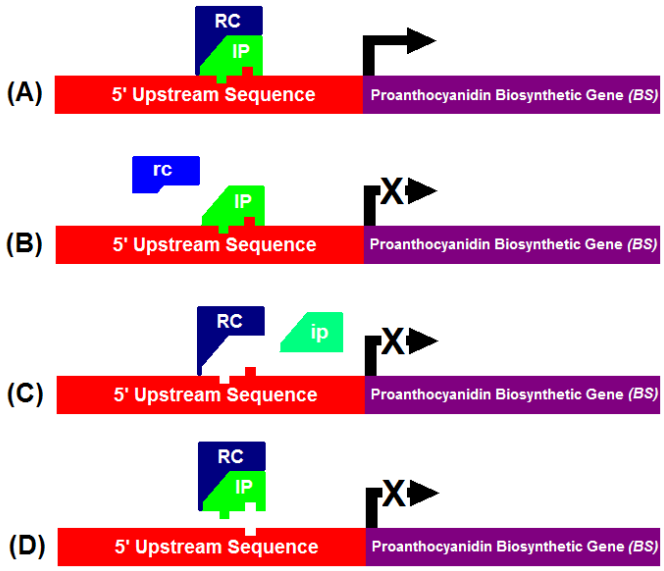


Figure 9.1: Model for RC-mediated transcriptional activation of a proanthocyanidin biosynthetic gene in the rice pericarp. This model is based on previously reported data that demonstrate an interaction between a bHLH protein (in this case, RC) and one or several interacting partner proteins (IP) is necessary for successful transcription of biosynthetic genes in the anthocyanin and proanthocyanidin pathways. In addition, specific recognition sequences (*cis*-elements) in the promoter of the biosynthetic genes are necessary for proper signaling and interaction between the transcription factor proteins and the DNA molecule. Therefore, when RC, IP, and the recognition sequence are intact (wild-type), the biosynthetic gene is expressed (A). Under this model, if RC (B), IP (C), or the recognition sequence in the promoter (D) are rendered nonfunctional, the biosynthetic gene is not expressed.

This model predicts that the functional RC protein cooperates with a functional interacting partner (IP) and these components directly interact with the upstream promoter region of a proanthocyanidin biosynthetic gene. When these three elements are present, the promoter sequence is activated, resulting in transcription of the biosynthetic gene (A). When any one of these elements is altered, so that the combinatorial interactions are disrupted, transcriptional activation is abolished (B, C, D).

Our hypothesis is that this RC/IP/promoter interaction imparts the specificity of proanthocyanidin accumulation in rice pericarp tissue. Our objectives are therefore twofold. First, we aim to identify a rice proanthocyanidin gene that is regulated by RC. This will require the expression of known flavonoid biosynthetic genes to be systematically tested in young pericarp tissue of red- and white-seeded rice genotypes to determine at which steps RC is acting. Secondly, we will dissect the promoter of an RC-regulated proanthocyanidin biosynthetic gene to determine the region that is necessary for transcriptional activation by RC. This will involve a) making a set of at least five, contiguous 5' deletions in the upstream region of the target biosynthetic gene and fuse these deletions to the GUS or GFP reporter gene; b) transforming the constructs from a) into red- and white-seeded rice genotypes; and c) analyzing GUS or GFP expression in pericarp tissue to identify a sub-region of the promoter that is necessary for RC-mediated transcriptional activation.

EXPERIMENTAL APPROACHES AND PRELIMINARY RESULTS

Identifying rice proanthocyanidin biosynthetic gene(s) regulated by RC.

The *Rc* gene encodes a bHLH protein that regulates the proanthocyanidin biosynthetic pathway in the pericarp tissue of rice (Sweeney et al., 2006). While proanthocyanidin biosynthesis in the grain is dependent upon the presence of a functional *Rc* (demonstrated by the fact that all white-grained genotypes have a defective *rc* allele), the step in the pathway at which the RC protein acts has yet to be determined. The first goal of this project was therefore to determine which biosynthetic gene(s) related to proanthocyanidin synthesis are affected directly or indirectly by the presence of a functional RC protein. Once this is known, we will focus on dissecting the promoter sequence of a target biosynthetic gene (*BS*) that is transcriptionally regulated by RC.

Different families of flavonoids all share a common set of initial steps leading from naringenin chalcone to dihydroflavonols (Figure 9.2). This is followed by the production of leucoanthocyanins by an NADPH-dependent dihydroflavonol reductase (DFR) enzyme. From here, the proanthocyanidin pathway branches off from the anthocyanin pathway through conversion of leucoanthocyanins to catechins by an NADPH-dependent leucoanthocyanin reductase (LAR). Proanthocyanidin precursors are transported into the vacuole by a *Bronze*-like vacuolar transporter. Finally, a yet unknown condensing enzyme performs the final step by linking catechin and leucoanthocyanin units to form proanthocyanidin polymers.

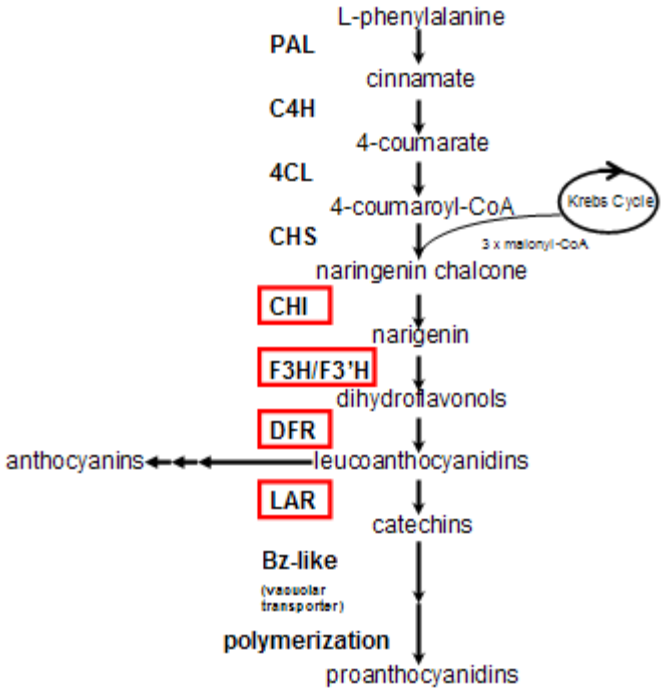


Figure 9.2: The general proanthocyanidin biosynthetic pathway in plants. The biosynthesis of proanthocyanidins in plants is part of the general phenylpropanoid pathway that begins with the amino acid phenylalanine. Red boxes: The expression of an early biosynthetic gene (CHI) and several late biosynthetic genes (F3H, F3'H, DFR, LAR) was analyzed via RT-PCR (Figure 9.3).

Functional copies of chalcone synthase (CHS), chalcone isomerase (CHI), and dihydroflavonal-4-reductase (DFR) have previously been cloned from rice (Reddy et al., 1996; Nakai et al., 1998; Druka et al., 2003). We have identified rice homologs of flavonone-3-hydroxylase (F3H), flavonone-3'-hydroxylase (F3'H), leucoanthocyanin reductase (LAR), and the *Bronze*-like flavonol-3-*O*-glucosyltransferase (*Bz*-like) based on BLAST searches and complex database queries using template sequences from other model species.

In *Arabidopsis*, where proanthocyanidin biosynthesis and regulation have been extensively studied, the *TT8* gene, encoding a bHLH protein closely related to RC (Sweeney et al., 2006), is required for the expression of at least two late biosynthetic genes (LBGs), *DFR* (dihydroflavonol reductase) and *BAN* (leucoanthocyanidin reductase; *LAR*) (Nesi et al., 2000). RC-like proteins have also been found to regulate the expression of *DFR*-orthologous genes from other species, such as in snapdragon (DELILA) (Martin et al., 1991), petunia (AN1) (Beld et al., 1989), and maize (B) (Chandler et al., 1989), suggesting this mechanism of regulation is conserved among plants.

We analyzed the expression of each of these genes (red boxes, Figure 9.2) by extracting total RNA from excised young pericarp tissue of red- and white-seeded rice genotypes (as described in (Sweeney et al., 2006)). We chose white-seeded genotypes that exhibited purple or red pigmentation in vegetative tissues, indicating that the full suite of flavonoid biosynthetic enzymes was present and functional in the plant. RNA was isolated using Qiagen RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) and was subsequently treated with DNase I enzyme (Invitrogen, Carlsbad, CA, USA) to remove residual DNA. cDNA synthesis was then performed using the SuperScript III First Strand Synthesis Kit (Invitrogen) according to manufacturer protocols.

The expression patterns of the targeted proanthocyanidin biosynthetic genes between red- (*Rc*) and white- (*rc*) seeded rice genotypes were systematically examined using RT-PCR. Our expectation was that some of these genes will be expressed in the pericarp tissues of red-seeded plants, but not in those of white-seeded plants. This would indicate what genes in the proanthocyanidin pathway RC is acting upon to

control red pigment accumulation in the rice grain. The results of this RT-PCR survey are illustrated in Figure 9.3.

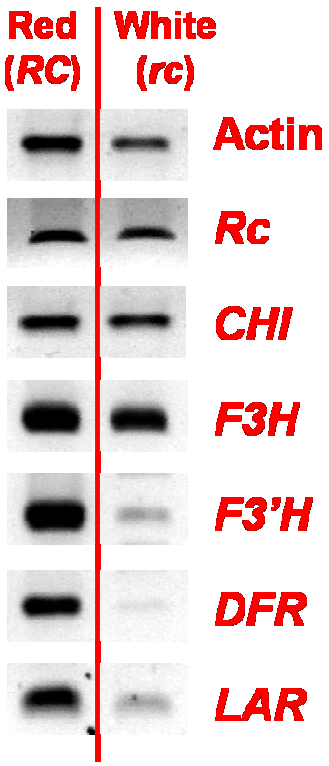


Figure 9.3: RT-PCR results for proanthocyanidin biosynthetic genes in rice pericarp. Total RNA was extracted from excised young pericarp tissue from a rice variety having red pericarp (*Rc*) and white pericarp (*rc*). Following normalization and cDNA synthesis, RT-PCR was performed for each of the genes with a red box in Figure 9.2. Actin1 was used as a standard. Note that while *Rc* is defective in the *rc* mutants, the expression of *Rc* does not change. The expression of *CHI* and *F3H* appear unchanged in the *rc* mutant, while *F3'H*, *DFR*, and *LAR* have reduced expression in the *rc* mutant.

These results demonstrated that the early biosynthetic genes did not appear to be affected by the presence or absence of a functional RC protein, which is consistent with the results of similar gene expression surveys in *Arabidopsis* (Nesi et al., 2000). In contrast, several of the late biosynthetic genes, including *F3'H*, *DFR* and *LAR*,

showed a marked reduction in expression in the *rc* mutant. This suggests that the RC protein potentially acts on several of the late proanthocyanidin biosynthetic genes.

The *DFR* gene on rice chromosome 1 showed negligible expression in the *rc* mutant background, making it a candidate for continued analysis. This *DFR* gene corresponds to the classically mapped *Rd* gene (Kato and Ishikawa, 1921), which was recently cloned and characterized (Furukawa et al., 2007). Given the apparent epistatic relationship between *Rc* and *Rd* in rice (see Introduction), and given the results from *Arabidopsis* demonstrating that *DFR* is regulated by bHLH proteins (which condition seed coat pigmentation), we believe this *DFR* gene is regulated by RC. The promoter region of this gene was therefore targeted for further analysis to more precisely determine the DNA sequence necessary for RC-mediated regulation.

Dissection of the 5' region of *DFR* to determine the promoter region that is necessary for transcriptional activation by RC.

In order to expand our knowledge of the promoter regions key to proanthocyanidin regulation in the rice pericarp, we analyzed the promoter region of *DFR*. This knowledge will also be essential for the development of an efficient and highly specific transgene containment strategy for rice. The extensive work done on the flavonoid pathway in other plant systems provided us with a framework for targeting our studies of flavonoid biosynthetic gene promoters. Our model predicts the binding of RC, its interacting partner, or both to the *DFR* promoter are required for proper *trans*-activation (Figure 9.1).

Since the published rice genome sequence is based on cv. Nipponbare, which has white pericarp (due to being an *rc* mutant, but also potentially accompanied by a

defective *DFR*), we first sequenced a ~2 kb region upstream of *DFR* in a red genotype to ensure we were analyzing a functional promoter. Then, using previously reported data and publicly available *cis*-element databases (such as PLACE: Plant cis-acting DNA elements), we defined candidate regions (bHLH and Myb binding motifs) of the *DFR* promoter as the basis for a targeted deletion analysis.

To create the promoter deletions, we first PCR amplified a region from a red genotype (RA 5755; *tropical japonica*) approximately 2 kb 5' from the *DFR* transcriptional start site. The amplification product was cloned into TOPO TA vector (Invitrogen) and the insert was sequenced to ensure no errors were introduced during amplification. Serial deletions were then constructed by PCR amplification using internal primers (Supplemental Table 9.1) to amplify deletion products at intervals that target putative *cis*-elements (as described by (Tyagi et al., 2005)) (Figure 9.4). All deletion inserts were initially cloned into a modified plant transformation vector, pCAMBIA1301 (CAMBIA, Canberra, Australia). The pCAMBIA1301 vector was modified by removing the 35S promoter from upstream of the GUS reporter gene, allowing ligation of the *DFR* promoter deletions upstream of GUS (Figure 9.5). The restriction enzyme sites used to remove the 35S promoter were HindIII (located in the pUC18 multiple cloning site (MCS)) and NcoI (located just after the 35S promoter). Later, the inserts were also cloned into a modified pCAMBIA1302_mGFP5 plant transformation vector, where the *DFR* promoter deletions would drive the expression of the GFP reporter gene (Figure 9.6). The same restriction enzyme sites were used for subcloning into the pCAMBIA1302 plasmid. To facilitate cloning of the *DFR* promoter deletions into these transformation vectors, HindIII and NcoI recognition sequences were added to the 5' (forward primer) or 3' (reverse primer) ends of the PCR primers used to amplify each deletion. Following PCR amplification of each

deletion from the TOPO-TA vector containing the 2 kb promoter fragment, the amplified products were double digested with HindIII and NcoI. The pCAMBIA vectors were also double digested with these enzymes, providing "sticky ends" and ensuring proper directional cloning of the promoter deletion inserts. Ligated constructs were transformed into *E. coli* dh5α competent cells, which were allowed to grow for 1 hour in SOC media in the absence of kanamycin, and then plated on kan+ agar plates (75µl/ml). Following overnight incubation, colony-PCR was performed to identify successfully transformed colonies. PCR was performed using a primer common to all the *DFR* deletion inserts (Del5_F) and a reverse primer in the vector backbone (NOS_R). Positive colonies were grown in kan+ liquid LB media overnight for mini-preps. Plasmid DNA was obtained using a Qiagen Mini-Prep kit. Excess bacterial culture was used to prepare glycerol stocks for each *DFR* promoter deletion, which were stored in the -80C freezer. Confirmation of each construct was performed by both digestion with HindIII and NcoI to observe insert fallout and by PCR with a primers anchored in the vector (Supplemental Table 9.2).

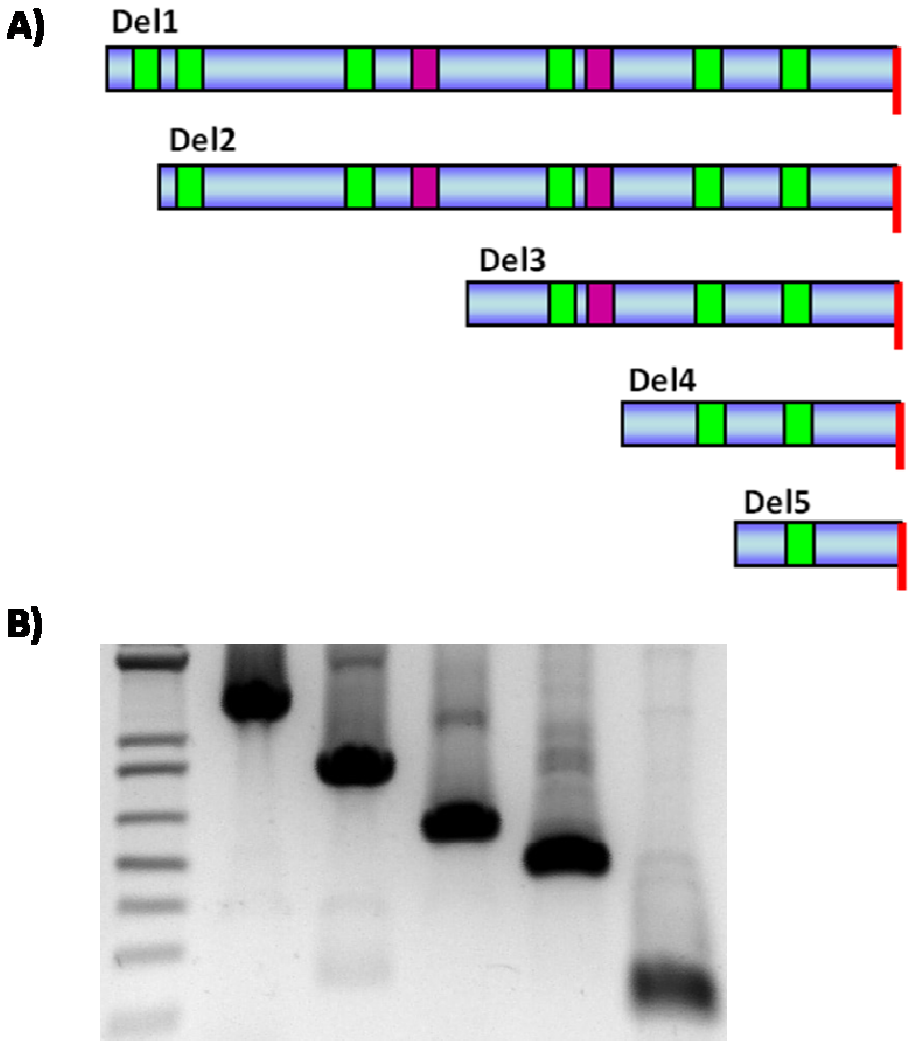


Figure 9.4: DFR Promoter Deletions. A 2 kb region 5' of the *DFR* gene was amplified from an accession having red pericarp (RA 5755). This fragment was cloned into a TOPO-TA vector.

A) Five promoter deletions (Del1–5) were created from this large fragment, with each deletion targeting the successive removal of putative *cis*-regulatory elements involved in proanthocyanidin biosynthesis. Putative bHLH- and myb-binding sequences are illustrated as green and purple boxes, respectively. The red line indicates the translation initiation codon for *DFR*.

B) Agarose gel picture of each DFR promoter deletion. Del1= 1203 bp; Del2= 802 bp; Del3= 546 bp; Del4= 422 bp; Del5= 151 bp.

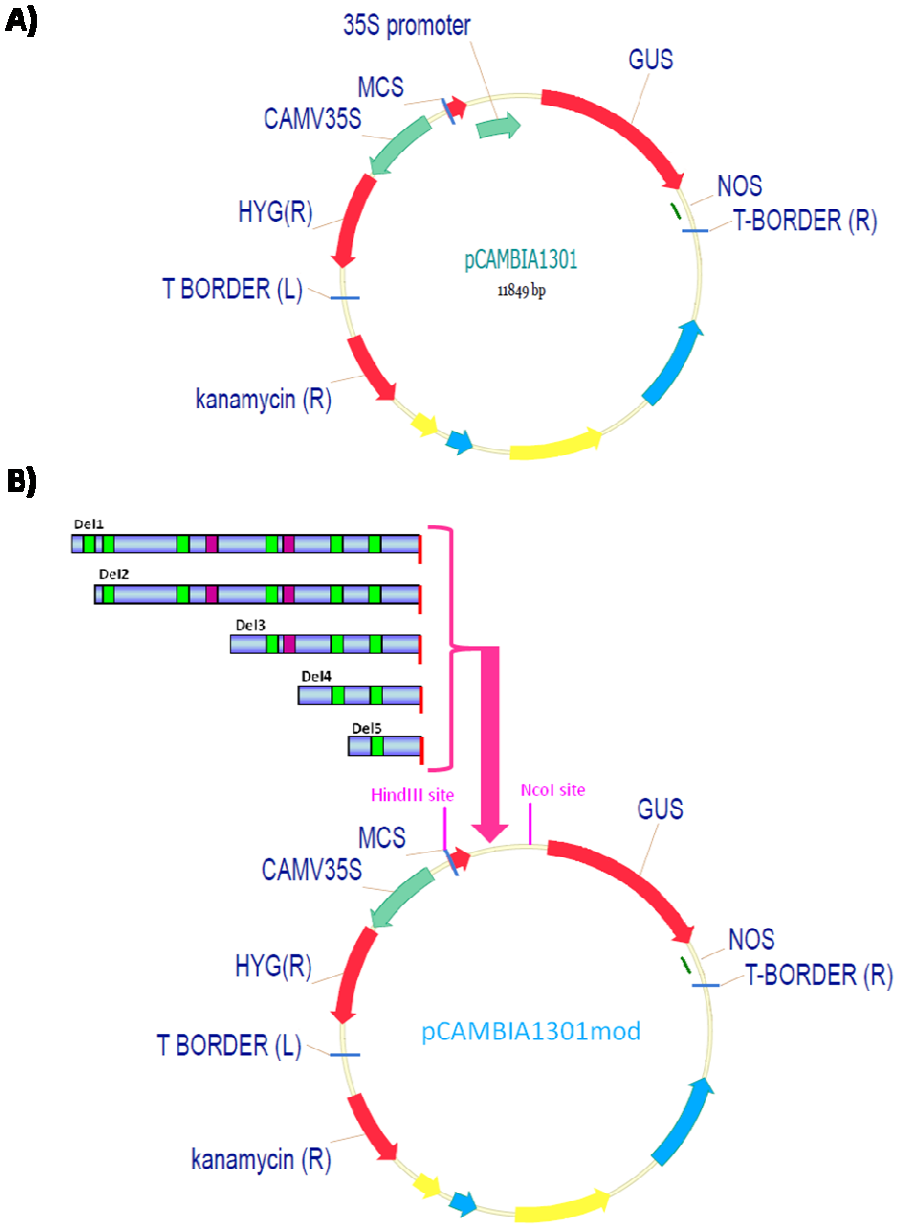


Figure 9.5: pCAMBIA1301 and pCAMBIA1301mod

A) Original pCAMBIA1301 vector. In the native vector, the CMV 35S promoter drives expression of the GUS reporter gene.
B) Modified pCAMBIA1301. The 35S promoter was removed by double digesting the plasmid with HindIII and NcoI. Each *DFR* promoter deletion was then cloned into the vector to drive the expression of GUS.

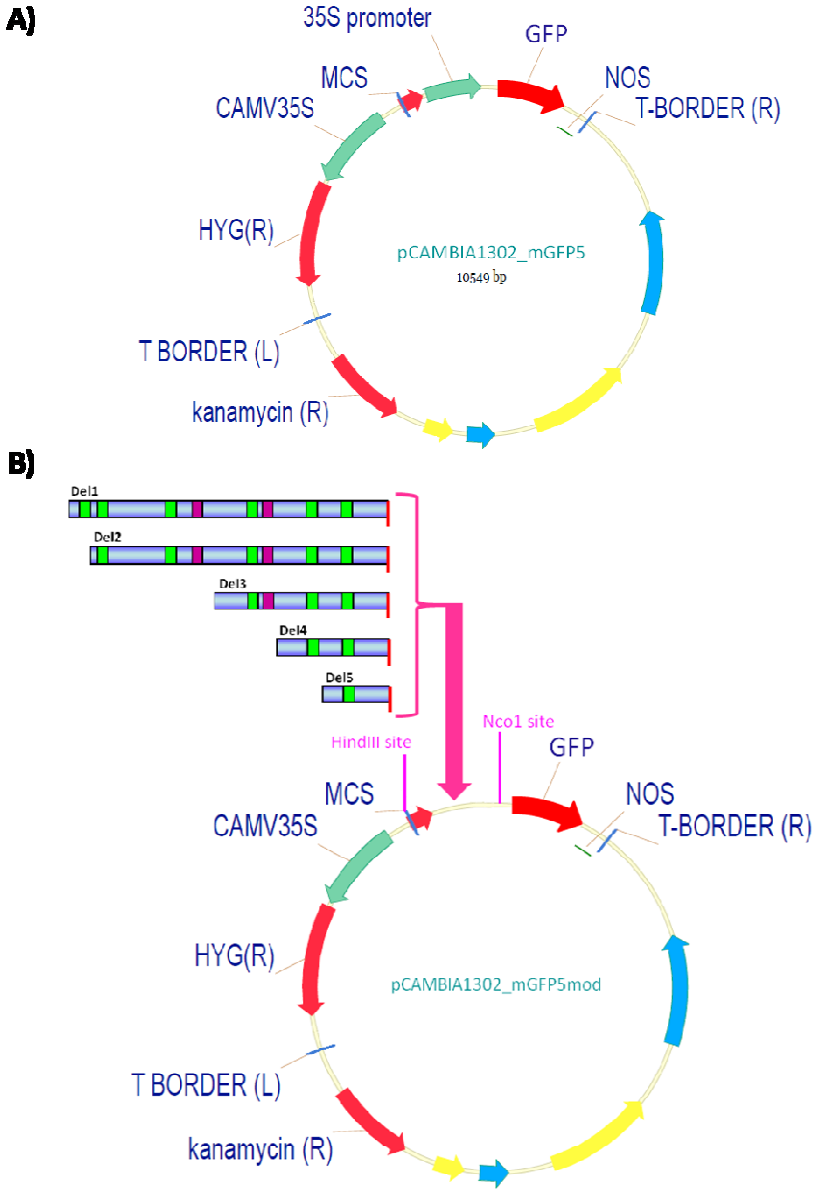


Figure 9.6: pCAMBIA1302_mGFP and pCAMBIA1302_mGFP5mod

A) pCAMBIA1302_mGFP5 vector. In the vector first obtained for this project, the CMV 35S promoter drives expression of the GFP reporter gene.
B) Modified pCAMBIA1302_mGFP5. The 35S promoter was removed by double digesting the plasmid with HindIII and NcoI. Each *DFR* promoter deletion was then cloned into the vector to drive the expression of GFP.

These deletion constructs were initially used in a series of transient assay experiments whereby the constructs are introduced into excised young pericarp tissue via particle bombardment. Transient assay experiments were carried out on varieties having both red and white pericarp. The red genotype has an active and functional RC protein that presumably must be present to activate the promoter constructs, while the white-seeded genotype will serve as a negative control, because an RC-responsive promoter should not be expressed in the absence of a functional RC protein. When these initial transient assays were attempted, GUS expression was observed in pericarp tissue from both the red and white variety, likely indicating "leaky" GUS expression using this technique. Given these unsuccessful results with the transient assay, we decided to create stable transformants with each *DFR* promoter deletion construct.

Due to the greater amenability of *Japonica* genotypes to transformation, we selected red and white *Japonica* genotypes for our transformation experiments. This transformation work has been carried out by Dr. Yong-Gu Cho at Chungbuk National University, Cheongju, Republic of Korea. The transformation procedures followed a modified method described previously (Toki et al., 2006). *Agrobacterium* strain EHA105 was transformed with the five pCAMBIA1301 constructs (Del1, 2, 3, 4, 5) and five pCAMBIA1302 constructs (Del1, 2, 3, 4, 5). Mature rice seeds of *Japonica* rice cultivar Gopumbyeo (white) and Jeokjinjubyeo (red) were dehulled and healthy seeds were selected. Dehulled seeds were sterilized with 70% ethanol 1 min prior to washing in sterile water. Seeds were further sterilized with 2.5% sodium hypochlorite containing 1 drop of Tween 20 per 50 ml for 15 min then washed five times in sterile water. This step was repeated once without Tween 20. The sterilized seeds were inoculated on N6D medium solidified with 0.4% Gelrite and cultured under continuous light at 32°C for 5 days. *Agrobacterium* strain EHA105 harboring either

binary Ti plasmid pCAMBIA1301 or pCAMBIA1302 containing the *DFR* promoter deletions was cultured on AB medium containing 50 mg/L kanamycin sulfate solidified with 1.5% agar for 3 days at 28°C in the dark. *Agrobacterium* culture was scraped from the plates and suspended in AAM medium to yield an OD_{600} of approximately 0.1. Pre-cultured seeds were immersed in the *Agrobacterium* suspension by gently inverting the tube for 1.5 min and then blotted dry with a sterilized filter paper to remove excess bacteria. These seeds were transferred onto a sterilized filter paper that had been moistened with 0.5 ml of AAM medium placed on 2N6-AS medium solidified with 0.4% Gelrite. After 3 days of co-cultivation at 25°C in the dark, seeds were washed five times in sterilized water and then washed once in sterilized water containing 500 mg/L carbenicillin to remove *Agrobacterium*. The seeds were rapidly blotted dry on a sterilized filter paper and cultured on N6D medium containing 50 mg/L hygromycin and 400 mg/L carbenicillin under continuous light at 32°C for 2 weeks. Proliferating calli arising from the scutellum were transferred to RE-III medium. Plantlets arising from the calli were transferred to HF medium to induce roots. Plantlets were transferred to soil 2-3 months after initiation of seed culture. This transformation experiment was expected to take 6-8 months from seed-to-seed. Integration of the insert was confirmed first by genomic PCR and then by Southern analysis. This experiment will allow us to assay for GUS/GFP activity in stable transgenic rice plants to determine the minimal *DFR* promoter necessary for activation in the presence of an active RC protein. As of January 2010, stable transformants were created and confirmed for several of the *DFR* promoter deletion constructs, although only T_0 plants (hemizygous) were being cultivated. Future GUS/GFP screening will need to be done on grain from T_1 plants that are homozygous for the *DFR* promoter/reporter insertion.

We will either perform histochemical assays to detect GUS activity or use confocal microscopy to analyze GFP activity in intact seeds or excised young pericarp tissue of transgenic plants, compared to control tissues. It is necessary to assay for GUS/GFP to confirm the activity of the deletion constructs in a red-seeded background, where the native *DFR* promoter is also active. We expect that one or more of our promoter deletion constructs will result in a loss of GUS/GFP expression, allowing us to conclude that the deleted portion of the promoter is essential for RC-mediated expression.

The above experiment assumes that the functionally important region of the *DFR* promoter is within 2 kb upstream of the transcriptional start site of the gene. It is possible that the important regulatory elements may be further upstream or elsewhere in the genome. However, work on flavonoid pathway regulation in maize and *Arabidopsis* has revealed that a small region of the promoter near the transcription start site contains putative Myb and bHLH binding sites and that this region is necessary and sufficient for tissue-specific activation of several flavonoid biosynthetic genes (Roth et al., 1991; Tuerck and Fromm, 1994; Bodeau and Walbot, 1996; Lesnick and Chandler, 1998; Debeaujon et al., 2003). If needed, we will extend our experiments to include regions farther upstream in the promoter. Another potential limitation with this experiment is if *DFR* is still expressed at low levels in the *rc* mutant (which the RT-PCR results would seem to suggest (Figure 9.3)), which would confound our results.

These promoter deletion analyses aim to establish the region of the *DFR* promoter that is necessary and required for efficient expression of the gene. This promoter sequence will be the target for our future efforts to develop a transgene containment strategy in

rice. The "contained" construct would express an embryo-lethal gene only in genotypes and tissues having a functional RC protein, preventing the survival of red-seeded transgenic genotypes, and therefore providing protection against transgene contamination of the wild species.

Conclusions and Future Directions

The use of rice in our studies allows us to utilize the resources available in this model plant species and then apply our results to a crop where this pathway has immediate importance to U.S. and world agriculture. These studies will lay the foundation for a more in-depth characterization of the proanthocyanidin pathway in rice pericarp, including the determination of the *cis*-elements required for RC-mediated activation. By dissecting the mechanisms of RC-mediated transcriptional control, the research put forth in this proposal will also help address a fundamental problem in plant biology: to define the molecular mechanisms by which bHLH proteins, one of the largest families of transcription factors in plants, mediate transcriptional regulation of gene expression. In addition, the flavonoid pathway represents an interesting model for the study of metabolic and regulatory networks, providing insight into secondary metabolism in other plant systems. For example, knowledge of the proanthocyanidin pathway in rice is of interest to the wheat community, where proanthocyanidins are responsible for economically important traits, including brightness of wheat flour and level of grain dormancy (Himi et al., 2005). Ultimately, this work will lead to the creation of a new transgene containment strategy in rice that will protect against the transmission of transgenes to wild and weedy rice populations. By creating a construct that responds differentially to the presence of a specific "wild" versus "cultivated" allele, domesticated gene pools can be maintained in isolation from wild or weedy gene pools, protecting both from unwanted gene flow. This concept could be applied more

generally to diverse transgenic varieties, alleviating some of the most salient fears related to transgene contamination of wild germplasm. To provide a basis for this preliminary research toward the development of a "contained" transgene construct, a Patent Cooperation Treaty (PCT) has been filed with the World Intellectual Property Organization (WIPO). The genetic diversity contained within the wild populations of rice is of vital importance to future generations of rice breeders, and our goal is to ensure these resources are available for utilization far into the future.

Supplemental Table 9.1: DFR Promoter Deletion Primers

Name	Primer Sequence
DFR_DEL1_F	CCCAAGCTTCAACGACGACCAGGTTCAAAT
DFR_DEL2_F	CCCAAGCTTCTGGAGGAGCTACAAGGAGAGG
DFR_DEL3_F	CCCAAGCTTCTGTCTCCTTTGCCTGTCGT
DFR_DEL4_F	CCCAAGCTTCGCTGGTCATTCTGTCTACTCC
DFR_DEL5_F	CCCAAGCTTCCCTGTCCTGTACCAGCTTA
DFR_DEL_R	CATGCCATGGTTCGCACGCGAATCTAACAC

*Note: HindIII restriction enzyme sites have been added to the 5' end of each forward primer and an NcoI site is added to the 3' end of the common reverse primer.

Supplemental Table 9.2: Other Useful Primers for Cloning

Name	Primer Sequence
GUS_F	ACTCGTCCGTCCTGTAGAAAC
GUS_R	CCAGTTCAGTTCGTTGTTCAC
GFP_R	TCACCCTCTCCACTGACAGA
pUC18_F_1	GGTACCCGGGGATCCTCTA
NOS_R_1	CAAGACCGGCAACAGGATTC

REFERENCES

Atchley W.R., Terhalle W., Dress A. (1999) Positional dependence, cliques, and predictive motifs in the bHLH protein domain. Journal of Molecular Evolution 48:501-516

Atkins J.F., Weiss R.B., Thompson S., Gesteland R.F. (1991) Genetic and developmental control of anthocyanin biosynthesis. Annual Review of Genetics 25:173-199

Beld M., Martin C., Huits H., Stuitje A.R., Gerats A.G.M. (1989) Flavonoid synthesis in *Petunia hybrida*: partial characterization of dihydroflavonol-4-reductase genes. Plant Molecular Biology 13:491-502

Bessey C.E. (1906) Crop improvement by utilizing wild species. Journal of Heredity os-2:112-118

Bharti A.K., Khurana J.P. (2003) Molecular characterization of transparent testa (tt) mutants of *Arabidopsis thaliana* (ecotype Estland) impaired in flavonoid biosynthetic pathway. Plant Science 165:1321-1332

Bodeau J.P., Walbot V. (1996) Structure and regulation of the maize *Bronze2* promoter. Plant Molecular Biology 32:599-609

Bovy A., de Vos R., Kemper M., Schijlen E., Almenar Pertejo M., Muir S., Collins G., Robinson S., Verhoeyen M., Hughes S., Santos-Buelga C., van Tunen A. (2002) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes *LC* and *C1*. Plant Cell 14:2509-2526

Brownlie P., Ceska T.A., Lamers M., Romier C., Stier G., Teo H., Suck D. (1997) The crystal structure of an intact human Max DNA complex: new insights into mechanisms of transcriptional control. Structure 5:509-520

Burgos N.R., Norman R.J., Gealy D.R. (2000) Comparative N uptake between red rice and rice. Proc South Weed Sci Soc 53:149

Burgos N.R., Shivrain V.K., Anders M.M., Sparks O.C., Rajuru S.N. (2003) Hybridization between Clearfield® rice and red rice under field conditions. AAES Research Series, pp 103-109

Burr F.A., Burr B., Scheffler B.E., Blewitt M., Wienand U., Matz E.C. (1996) The maize repressor-like gene *intensifier1* shares homology with the *r1/b1* multigene family of transcription factors and exhibits missplicing. Plant Cell 8:1249-1259

Carey C.C., Strahle J.T., Selinger D.A., Chandler V.L. (2004) Mutations in the *pale aleurone color1* regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar *TRANSPARENT TESTA GLABRA1* gene in *Arabidopsis thaliana*. Plant Cell 16:450-464

Chandler V.L., Radicella J.P., Robbins T.P., Chen J., Turks D. (1989) Two regulatory genes of the maize anthocyanin pathway are homologous: Isolation of *B* Uutilizing *R* genomic sequences. Plant Cell 1:1175-1183

Coffman R.W., Herrera R.M. (1980) In: Fehr W.R., Hadley H.H. (eds) Hybridization of Crop Plants. The American Society of Agronomy, Inc.

de Pater S., Pham K., Memelink J., Kijne J. (1997) RAP-1 is an *Arabidopsis* MYC-like R protein homologue, that binds to G-box sequence motifs. Plant Molecular Biology 34:169-174

de Vetten N., Quattrocchio F., Mol J., Koes R. (1997) The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes & Development 11:1422-1434

Debeaujon I., Nesi N., Perez P., Devic M., Grandjean O., Caboche M., Lepiniec L. (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: Regulation of differentiation and role in seed development. Plant Cell 15:2514-2531

Diarra A., Smith R.J., Jr., Talbert R.E. (1985) Growth and morphological characteristics of red rice (*Oryza sativa*) biotypes. Weed Science 33:310-314

Druka A., Kudrna D., Rostoks N., Brueggeman R., von Wettstein D., Kleinhofs A. (2003) Chalcone isomerase gene from rice (*Oryza sativa*) and barley (*Hordeum vulgare*): physical, genetic and mutation mapping. Gene 302:171-178

Duffy R. (2001) Specialty rices of the world: breeding, production and marketing. Science Publishers, Inc., Enfield, NH, USA

Ellenberger T., Fass D., Arnaud M., Harrison S.C. (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes & Development 8:970-980

Erguiza A., Duff B., Khan C. (1990) Choice of rice crop establishment techniques: Transplanting vs. wet seeded. International Rice Research Institute, Los Banos, Laguna, Philippines

FAO (2009) FAOSTAT Database, Crop Production. Dec. 16, 2009 edn. Food and Agricultural Organization of the United Nations

Furukawa T., Maekawa M., Oki T., Suda I., Iida S., Shimada H., Takamure I., Kadowaki K.-i. (2007) The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. The Plant Journal 49:91-102

Gealy D.R., Mitten D.H., Rutger J.N. (2003) Gene flow between red rice (*Oryza sativa*) and herbicide-resistant rice (*O. sativa*): Implications for weed management. Weed Technology 17:627-645

Gealy D.R., Tai T.H., Sneller C.H. (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Science 50:333-339

Goff S.A., Cone K.C., Chandler V.L. (1992) Functional analysis of the transcriptional activator encoded by the maize *B* gene: evidence for a direct functional interaction between two classes of regulatory proteins. Genes & Development 6:864-875

Gong Z.-Z., Yamagishi E., Yamazaki M., Saito K. (1999) A constitutively expressed Myc-like gene involved in anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in transgenic plants and transactivation in yeast cells. Plant Molecular Biology 41:33-44

Goss W.L., Brown E. (1939) Buried red rice. Journal of the American Society of Agronomy 31

Grotewold E., Drummond B.J., Bowen B., Peterson T. (1994) The myb-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell 76:543-553

Grotewold E., Sainz M.B., Tagliani L., Hernandez J.M., Bowen B., Chandler V.L. (2000) Identification of the residues in the Myb domain of maize *C1* that specify the interaction with the bHLH cofactor *R*. Proceedings of the National Academy of Sciences of the United States of America 97:13579-13584

Hernandez J.M., Heine G.F., Irani N.G., Feller A., Kim M.-G., Matulnik T., Chandler V.L., Grotewold E. (2004) Different mechanisms participate in the R-dependent activity of the R2R3 MYB transcription factor *C1*. Journal of Biological Chemistry 279:48205-48213

Himi E., Nisar A., Noda K., Schwarzacher T. (2005) Colour genes (*R* and *Rc*) for grain and coleoptile upregulate flavonoid biosynthesis genes in wheat. Genome 48:747-754

Holton T.A., Cornish E.C. (1995) Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell 7:1071-1083

Kato S., Ishikawa J. (1921) On the inheritance of the pigment of red rice. Japanese Journal of Genetics 1:1-7

Kawagoe Y., Murai N. (1992) Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein *phaseolin* gene conferring spatial and temporal control. The Plant Journal 2:927-936

Kubo H., Peeters A.J.M., Aarts M.G.M., Pereira A., Koornneef M. (1999) *ANTHOCYANINLESS2*, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. Plant Cell 11:1217-1226

Kwon S.L., Smith R.J., Jr., Talbert R.E. (1992) Comparative growth and development of red rice (*Oryza sativa*) and rice (*O. sativa*). Weed Science 40:57-62

Lesnick M.L., Chandler V.L. (1998) Activation of the maize anthocyanin gene *a2* is mediated by an element conserved in many anthocyanin promoters. Plant Physiology 117:437-445

Ling W.H., Cheng Q.X., Ma J., Wang T. (2001) Red and black rice decrease atherosclerotic plaque formation and increase antioxidant status in rabbits. Journal of Nutrition 131:1421-1426

Lloyd A., Walbot V., Davis R. (1992) *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science 258:1773-1775

Loulergue C., Lebrun M., Briat J.-F. (1998) Expression cloning in Fe2+ transport defective yeast of a novel maize MYC transcription factor. Gene 225:47-57

MacKown C.T., Carver B.F., Edwards J.T. (2008) Occurrence of condensed tannins in wheat and feasibility for reducing pasture bloat. Crop Science 48:2470-2480

Maclean J.L., Dawe D.C., Hardy B., Hettel G.P., (eds). (2002) Rice Almanac 3rd Edition,

edn. CABI Publishing, Oxon, UK

Martin C., Prescott A., Mackay S., Bartlett J., Vrijlandt E. (1991) Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. The Plant Journal 1:37-49

McCouch S. (2004) Diversifying selection in plant breeding. PLoS Biology 2:e347

Moody K. (1992) Weed management in wet-seeded rice in tropical Asia. Food and Fertilizer Technology Center for the Asian and Pacific Region, Los Baños, Laguna, Philippines.

Mooney M., Desnos T., Harrison K., Jones J., Carpenter R., Coen E. (1995) Altered regulation of tomato and tobacco pigmentation genes caused by the *delila* gene of *Antirrhinum*. The Plant Journal 7:333-339

Nakai K., Inagaki Y., Nagata H., Miykzaki C., Iida S. (1998) Molecular characterization of the gene for dihydroflavonol-4-reductase of *japonica* rice varieties. Plant Biotechnology 15:221-225

Nesi N., Debeaujon I., Jond C., Pelletier G., Caboche M., Lepiniec L. (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. Plant Cell 12:1863-1878

Oki T., Masuda M., Kobayashi M., Nishiba Y., Furuta S., Suda I., Sato T. (2002) Polymeric procyanidins as radical-scavenging components in red-hulled rice. Journal of Agricultural and Food Chemistry 50:7524-7529

Quattrocchio F., Wing J.F., Leppen H., Mol J., Koes R.E. (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell 5:1497-1512

Quattrocchio F., Wing J.F., va K., de, Woude, Mol J.N.M., Koes R. (1998) Analysis of bHLH and MYB domain proteins: species-specific regulatory differences

are caused by divergent evolution of target anthocyanin genes. The Plant Journal 13:475-488

Ray H., Yu M., Auser P., Blahut-Beatty L., McKersie B., Bowley S., Westcott N., Coulman B., Lloyd A., Gruber M.Y. (2003) Expression of anthocyanins and proanthocyanidins after rransformation of alfalfa with maize *Lc*. Plant Physiology 132:1448-1463

Reddy A.R., Scheffler B., Madhuri G., Srivastava M.N., Kumar A., Sathyanarayanan P.V., Nair S., Mohan M. (1996) Chalcone synthase in rice (*Oryza sativa* L.): Detection of the CHS protein in seedlings and molecular mapping of the *chs* locus. Plant Molecular Biology 32:735-743

Roth B.A., Goff S.A., Klein T.M., Fromm M.E. (1991) *C1*- and *R*-dependent expression of the maize *Bz1* gene requires sequences with homology to mammalian myb and myc binding sites. Plant Cell 3:317-325

Sagasser M., Lu G.-H., Hahlbrock K., Weisshaar B. (2002) *A. thaliana TRANSPARENT TESTA 1* is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes & Development 16:138-149

Sainz M.B., Grotewold E., Chandler V.L. (1997) Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related myb domain proteins. Plant Cell 9:611-625

Sanders D.E., Strahan R.E., Linscombe S.D., Croughan T.P. (1998) Control of red rice (*Oryza sativa*) in imidazolinone tolerant rice. Proc South Weed Sci Soc 51:36-37

Serrano F.B. (1975) Twenty-six years of rice research and development. National Research Council of the Philippines, Quezon City, Philippines

Sompornpailin K., Makita Y., Yamazaki M., Saito K. (2002) A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens*. Plant Molecular Biology 50:485-495

Spelt C., Quattrocchio F., Mol J.N.M., Koes R. (2000) *anthocyanin1* of *Petunia* encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell 12:1619-1632

Sweeney M.T., Thomson M.J., Pfeil B.E., McCouch S. (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294

Tanksley S.D., McCouch S.R. (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066

Toki S., Hara N., Ono K., Onodera H., Tagiri A., Oka S., Tanaka H. (2006) Early infection of scutellum tissue with *Agrobacterium* allows high-speed transformation of rice. The Plant Journal 47:969-976

Tuerck J.A., Fromm M.E. (1994) Elements of the maize *A1* promoter required for transactivation by the anthocyanin *B*/*C1* or phlobaphene *P* regulatory genes. Plant Cell 6:1655-1663

Tyagi W., Rajagopal D., Singla-Pareek S.L., Reddy M.K., Sopory S.K. (2005) Cloning and regulation of a stress-regulated *Pennisetum glaucum* vacuolar ATPase-c gene and characterization of its promoter that is expressed in shoot hairs and floral organs. Plant and Cell Physiology 46:1411-1422

USDA (2009) USDA Agricultural Projections to 2018. In: Agriculture U.S.D.o. (ed)

Webster T.M. (2000) The southern states 10 most common and troublesome weeds in rice. Proc South Weed Sci Soc 53:247-274

Winkel-Shirley B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology 126:485-493

Zamir D. (2001) Improving plant breeding with exotic genetic libraries. Nature Reviews Genetics 2:983-989

GLOSSARY

Panicle – a compound raceme; the inflorescence of a rice plant

Grain shattering – seed abscission; when ripe seed falls from the panicle before harvesting occurs

Dormancy – a physiological period of quiescence during which a mature seed will not germinate

Tiller – a branch of the rice plant including roots, culms, leaves and panicle (if it is a productive tiller)

Photoperiodic response – a plant's ability to flower in response to increasing or decreasing daylength

Outcrossing – the act of cross-pollination, where pollen from the flowers of one plant is distributed to another, promoting cross-hybridization

Inbreeding – the act of self-pollination, where pollen is contained within the flowers of a single plant, preventing cross-hybridization

Subpopulation – a population within a species that is genetically divergent from other populations within that species

Haplotype - a combination of alleles or sequence variants at multiple linked loci that are transmitted together

Isozymes – variants of an enzyme; proteins that differ in amino acid sequence, but catalyze the same reaction

Phylogenetic – referring to the evolutionary relationships between organisms

Retrotransposon – a type of transposon (mobile DNA) that moves through an RNA intermediate as it copies and pastes itself throughout an organism's genome

Molecular clock – using DNA polymorphisms between the sequences of two organisms to deduce the amount of time that has elapsed since they diverged from each other

Single nucleotide polymorphism (SNP) – a mutation involving the substitution of a single nucleotide that can be to distinguish the DNA of one organism from another

Introgression – the movement of a discrete portion of a genome from one genotype into another via meiotic recombination

Transcription factor – a protein that affects the transcription and thus expression of specific target genes

Abscission layer – a zone of cells at the interface of plant organs that breaks down during senescence, promoting the organ to fall off

Thresh – the process of removing ripe grains from harvested plants

Landrace – a genotype or heterogeneous mixture of genotypes selected and maintained by farmers for good performance in a particular geographical or ecological region

Pericarp – the outermost layer of cells in the seed coat

Cultivar – cultivated variety

Phenotype – the physical appearance of an organism

Genotype – the genetic makeup of an organism

Accession – a sample of seed collected to represent a species, population, or variety

Admixed – a rice plant whose genome contains DNA inherited from different subpopulations within the species as the result of outcrossing and recombination

Life history habit – the reproductive cycle of a plant; annual habit refers to plants that live, reproduce and die in one year or season, while perennial habit refers to plants whose life cycle spans more than one year

Apiculus – the tip of the lemma or palea; a small point at the tip of the seed hull

RFLP marker – restriction fragment length polymorphism; a DNA marker that detects genetic differences between individuals based on the presence or absence of restriction enzyme sites

SSR marker – simple sequence repeat; a DNA marker that detects genetic differences between individuals based on the length of short repetitive sequences found throughout the genome

F_{st} – a measure of population differentiation based on polymorphism data. It compares the genetic variability within and between populations. If two populations have completely divergent DNA sequences, their F_{st} value = 1, indicating complete differentiation. If two populations have identical DNA sequences, their F_{st} value = 0, indicating no differentiation. Therefore, when comparing two populations, the larger the F_{st} value (the closer to 1), the more differentiated those populations are from each other